Communicating Subjective Evaluations

Matthias Lang (LMU Munich)

Discussion Paper No. 120

October 11, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Communicating Subjective Evaluations*

Matthias Lang**

August 2018

Abstract

Consider managers evaluating their employees' performances. Should managers justify their subjective evaluations? Suppose a manager's evaluation is private information. Justifying her evaluation is costly but limits the principal's scope for distorting her evaluation of the employee. I show that the manager justifies her evaluation if and only if the employee's performance was poor. The justification assures the employee that the manager has not distorted the evaluation downwards. For good performance, however, the manager pays a constant high wage without justification. The empirical literature demonstrates that subjective evaluations are lenient and discriminate poorly between good performance levels. This pattern was attributed to biased managers. I show that these effects occur in optimal contracts without any biased behavior.

JEL classifications: D82, D86, J41, M52

Keywords: Communication, Justification, Subjective Evaluation, Centrality, Leniency, Disclosure

*Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged. I thank Helmut Bester, Yves Breitmoser, Christoph Engel, Leonardo Felli, Ludivine Garside, Thomas Giebe, Alia Gizatulina, Olga Gorelkina, Dominik Grafenhofer, Paul Heidhues, Martin Hellwig, Sergei Izmalkov, Sebastian Kodritsch, Johannes Koenen, Daniel Krähmer, Marco Ottaviani, Régis Renault, Ilya Segal, Caspar Siegert, Roland Strausz, Stefan Terstiege, three anonymous referees, and the editor for very helpful suggestions and discussions, and the audiences at the World Congress of the Game Theory Society, Econometric Society NASM and EM, CESifo Applied Micro, EEA annual meeting and seminars in Milan (Bocconi), Mannheim, Berlin, Bonn, Maastricht and Munich for comments.

**University of Munich (LMU), Department of Economics, Geschwister-Scholl-Platz 1, 80539 Munich (Germany), and CESifo, Munich (Germany), lang@uni-bonn.de

1 Introduction

This paper analyzes communication in a principal-agent model in which the principal's performance measure is unobservable to the agent and non-verifiable by third parties. Such subjective measures are widely used in practice because verifiable, i.e., objective, performance measures are often unavailable.[1] Their subjectivity allows the principal to choose whether and how to disclose and to justify her evaluation of the agent's work. Hence, a hold-up problem arises: The principal wants to incentivize the agent to exert work effort, but these incentives depend on an appropriate evaluation in the end. Therefore, HR departments and personnel policies place great emphasis on feedback and communication of evaluations.[2] Nevertheless, empirically, subjective evaluations are distorted, and wage dispersion for the best evaluations is low.[3] These empirical observations are referred to as biases, which supposedly arise from supervisors' mistakes. In response, a whole industry has sprung up to provide training for supervisors. Alternatively, "some companies go so far as to rate employees on a bell curve," requiring supervisors to match a given distribution with their evaluations, as forcefully advocated by, e.g., Jack Welch, a former and renowned CEO of General Electric (New York Times, 2013).[4] I show that both responses may be misplaced because adjusting wages and communication to the subjectivity of evaluations is optimal and does not require any bias. The wage and communication pattern results from optimal contracting with standard preferences.

To explain these empirical observations, I study the communication of subjective evaluations. In the model, an agent (he) works for a principal (she) who privately receives information about the agent's performance. The principal has two options: she directly reports or justifies her evaluation. Justification of subjective evaluations is a common HR practice: "92% require a review and feedback session as part of the appraisal process." (Dessler, 2008, p. 366) If the principal provides justification, she cannot send low messages for good performance. Nevertheless, the principal's message is not necessarily truthful, and providing justification is costly. Below, I discuss some interpretations of this type of justification. The agent replies with an unverifiable message as to whether justification was provided. As the messages are the only third-party enforceable information, the contract only depends on these messages. I investigate the resulting communication pattern: in equilibrium, the principal justifies only poor evaluations. In this case, wages increase in the evaluation. For good evaluations, the principal in equilibrium saves the trouble of explaining them and simply pays a high wage, which yields pooling and wage

[1] The extensive use of subjective performance measures is confirmed by Noe et al. (2018, p. 355), Gibbs et al. (2009), Porter et al. (2008, p. 148), Levin (2003), MacLeod and Parent (1999), and Murphy (1993). The reason is that agents can manipulate objective performance measures or multi-task problems. Consequently, Gibbons (1998, p. 120) concludes that "objective performance measures typically cannot be used to create ideal incentives."

[2] See, for example, Noe et al. (2018, Chapter 8), Dessler (2008, Chapter 9), or Porter et al. (2008, Chapter 8).

[3] Section 3.4 discusses these empirical observations in detail and provides references.

[4] *New York Times*, 2013 Nov. 24, Invasion of the annual reviews, Business News p.8. See also *Wall Street Journal*, 1999 Jun. 21, Raises and Praise or Out the Door: How GE's Chief Rates and Spurs His Employees, p. B1.

compression at the top: leniency —, i.e., agents receive the highest wages more often than the best performance occurs — and centrality —, i.e., variation in performance exceeds variation in wages at the top — arise endogenously from optimal contracting.

The intuition for this communication pattern is the following: First, it is never optimal to justify all evaluations because justification is costly. Second, if the agent is evaluated positively, he suspects no deviation by the principal because the principal pays higher wages for better evaluations. If the agent is evaluated negatively, the agent considers two possibilities: his performance was poor or the principal distorted her evaluation downwards to pay lower wages. To counter such suspicions by the agent, the principal justifies poor evaluations. Note that compared to common moral-hazard settings, additional incentives are necessary: ex-ante, the principal wants to justify poor evaluations ex-post. Ex-post, however, she wants to save on justification costs. The principal has no commitment power other than the contract. Consequently, she must design contractual terms that make it ex-post incentive-compatible for her to justify the evaluation. Finally, there is a clear intuition for centrality: by eliminating wage differences that would otherwise call for justification, the principal can save some justification costs. Remember that the agent cannot verify the evaluation without justification. Hence, instead of reporting an evaluation yielding higher payments, the principal would deviate and report an evaluation that does not require justification and yields lower payments. Therefore, no wage dispersion is feasible, and there is pooling with respect to wages if the principal provides no justification.[5] The resulting communication pattern is not limited to employment relations. It applies more generally to hold-up and moral hazard settings whenever the better-informed party can provide justification.

Technically, justifications are based on type-dependent message spaces for the principal. All messages are contractible but unverifiable. Moreover, only the agent observes whether the principal provides justification. Hence, a third party cannot tell whether a message is truthful or whether justification occurred. Justification limits the principal's scope for deviations in reporting her evaluation of the agent as low messages become unavailable for good performance. In a well-designed contract, the justification ensures that the principal cannot report poor evaluations for good performance. As an interpretation, consider a business analyst at a consulting firm who received praise from a client. Suppose the partners in the firm pretend that the analyst's work was poor and justify their assessment by saying that the client complained about the analyst. Although the business analyst is unaware of the partners' real evaluation of her work, she knows that the partners are lying making such assessments unavailable to the partners. Alternatively, we can interpret a poor evaluation that comes with justifications as verifying that the agent has failed a well-defined standard.

The remainder of this paper is organized as follows. Section 2 discusses the related literature.

[5] Murphy (1993, p. 49) summarizes the reasoning as follows: Principals have "nonpecuniary costs [here, justification costs] associated with performance appraisal, which leads them to prefer to assign uniform ratings rather than to carefully distinguish employees by their performance."

Section 3 sets up the model and characterizes the optimal communication pattern. Section 4 contains the concluding remarks. All proofs are relegated to the appendix.

2 Related Literature

There is an extensive body of literature on subjective performance measures. Usually, it is assumed that evaluations are observable and relationships are long-term. This yields implicit contracts — for example, in Goldluecke and Kranz (2013), Pearce and Stacchetti (1998), Compte (1998), Baker et al. (1994), and MacLeod and Malcomson (1989). Reputation effects created by the continuation value for both contracting parties allow subjective performance measures to gain credibility to be used as the basis for the agent's incentives. Li and Matouschek (2013) and Levin (2003) drop the assumption that subjective performance measures are perfectly observable by both contracting parties. In this case, optimal contracts often have a termination form, i.e., contracts end after observing poor performance. See also Malcomson (2012) and MacLeod (2007) for recent surveys. In contrast to these repeated interactions, subjective evaluations are also used in static settings.

MacLeod (2003) was the first to implement subjective performance measures in a static setting. He assumes that the agent receives a signal that is correlated with the principal's evaluation. Each party reports their information by simultaneously sending a public message. As the information structure is exogenously given, the principal cannot decide, depending on the performance measure, whether to justify her evaluation. Nonetheless, his results correspond to two special cases in my model. If the agent's and the principal's signals are correlated, MacLeod (2003) achieves the common second-best solution similar to obligatory or costless justification in my model in Proposition 1. If the signals are uncorrelated, the optimal contract in MacLeod (2003) resembles the case of prohibitively expensive justification in my model. Economically, the main difference between my paper and MacLeod (2003) is that I endogenize communication. This allows me to discuss the resulting communication pattern. Nevertheless, the case of imperfect correlation with a binding upper limit on third-party payments by MacLeod (2003) shares some features with my optimal contracts, but the reasoning and the proofs are different. First, I do not assume an upper limit on payments. Second, the agent receives no private signals about his performance. Instead, it is the principal's incentives – resulting from the contract and justification costs – to withhold and distort her evaluation that yields compression at the top. MacLeod and Tan (2017) extend the model of MacLeod (2003) by considering malfeasance and more general information structures between agent and principal, such as better-informed agents. In addition, they change the timing and study sequential messages with the agent or the principal sending their message first. In this dichotomy, I scrutinize authority contracts with the principal reporting first, although with justification as a different communication technology.

In the current paper, I follow a static approach. Some justification can be found in Fuchs

(2007), who considers a finitely repeated principal-agent model. He shows that it is optimal for the principal to announce her subjective evaluation only once at the end of the interaction. In this case, the agent does not learn whether good performance has already occurred. Hence, it is sufficient to penalize only the worst outcome while paying a constant wage following all other terminal histories. Brown and Heywood (2005) and Addison and Belfield (2008) provide additional justification for a static approach. They show empirically that performance evaluations are more likely to be used for employees with shorter expected tenure.

I also relate to the literature on endogenous contracts, such as Kvaløy and Olsen (2009). However, I do not assume any cost for writing specific contractual arrangements. The contract can be any functions of the messages, but justification is costly. Another paper discussing endogenous verifiability is Dewatripont and Tirole (2005). They allow both sides to exert effort to increase the probability of a verifiable message. Nonetheless, they do not consider moral hazard and subjective evaluations. Because justification allows partial verification of the performance measure, there is a parallel to the literature on costly state verification, such as Hart and Moore (1998), Gale and Hellwig (1985), and Townsend (1979). These models allow an investor to verify the firm's performance at a cost. They show the optimality of debt contracts, which are similar to optimal contracts in my paper because there is no verification for high payments. In this literature, however, contracts provide no incentives for the agent, and verification is contractible, so payments can depend on whether verification occurred, as in Townsend (1979), and contracts specify when to verify, as in Gale and Hellwig (1985). In my model, justification is not contractible. Hence, contracts cannot enforce justification directly, and payments cannot depend on whether justification was provided. The reason is that justification need not be truthful and cannot be verified directly by one of the contracting parties, while verification is truthful and verifiable. This is also why mixed strategies with respect to justification are not optimal in my setting, in contrast to, e.g., Townsend (1979).

To make deterministic verification strategies optimal, Krasa and Villamil (2000) dynamically extend costly state verification. An investor can verify the firm's performance at a cost. Nevertheless, the contract can be renegotiated after the firm learned the state of the world. Simple debt contracts are optimal in this setting, dominating stochastic contracts. In my model, there is commitment to a contract. Thus, renegotiations are impossible, but it must be sequentially optimal for the better-informed side to provide justification. Hence, equilibrium wages without justification must be higher than justified equilibrium wages. In Krasa and Villamil (2000), the less-informed investor verifies the firm's payments if these payments are below a threshold. This is another distinction between the literature on costly state verification and my paper. In my model, due to the nature of justification, the better-informed side chooses whether to provide justification. In contrast, the less-informed side usually chooses whether to verify. For example, Doornik (2010) applies costly state verification to moral hazard. She considers a setting where output is

contractible but private information of the principal. The principal offers the agent a payment. If the less-informed agent rejects the principal's offer, the output is verified at a cost to both sides, and the agent's wage is determined according to the realized output. Restricting contracts to use either the principal's offer or the verified output, the agent stochastically triggers verification for low wage offers. The agent is willing to verify because contracts specify higher wages following verification. This is feasible because contracts can make wages dependent on whether verification occurred, and, if verified, the state of the world. In my setting, there is no verifiable output, but contracts can use any available information, and there are no exogenous restrictions on the contracting space. In Doornik (2010), the agent triggers costly enforcement. Hence, optimal contracts are determined by two indifference conditions: The difference between the contractual payments and the settlement offer must make the agent indifferent between accepting and rejecting the principal's offer. In addition, the agent's rejection rate makes the principal indifferent between offering the high wage and the low wage. Both indifference conditions are absent from my analysis.

Furthermore, leniency and wage compression or centrality could also be caused by fairness or trust. According to Bernardin and Orban (1990, p. 197) "trust in appraisal accounted for a significant proportion of variance in performance ratings." In my model, justification in combination with the contractual terms establishes this trust. In Giebe and Gürtler (2012), Al-Najjar and Casadesus-Masanell (2001), and Rotemberg and Saloner (1993), this trust is created by the extent to which the principal's preferences incorporate the agent's well-being, in contrast to the standard preferences in my model.

Several papers consider different rationales for subjective evaluations, namely, as a signal about the agent's productivity. If the agent does not know her productivity, she can infer her productivity from the evaluation by the principal. Fuchs (2015) shows that if the principal pays a discretionary bonus for positive evaluations, the bonus payment makes her evaluation credible and allows for a separating equilibrium. Zábojník (2014) uses subjective evaluations to fine-tune the agent's effort choice in a multi-task setting. Subjective evaluations can supplement imperfect objective measures and allow the agent to infer her productivity level.[6] Suvorov and van de Ven (2009) analyze subjective evaluations as a signal about the agent's productivity if the agent is intrinsically motivated.

3 Evaluating the Agent's Work

3.1 Subjective Evaluations at Work

For motivation, I begin with a brief case study. Consider performance evaluations at Arrow Electronics, a Fortune 500 company, as documented in Hall and Madigan (2000). Employees

[6] See also Kambe (2006) for the combination of subjective evaluations and objective performance measures.

- In period 0, the principal proposes a contract to the agent.
- In period 1, the agent can accept the contract offer and determines effort e.
- In period 2, the principal observes the agent's performance p.
- In period 3, the principal either reports evaluation $m_P \in \mathcal{P}$ or provides justification $m_P \in \{m \in \mathcal{P} | m \geq p\}$ at cost κ.
- In period 4, the agent sends a message $m_A \in \{0, 1\}$.
- In period 5, the principal makes payments according to $W(m_P, m_A)$ and $C(m_P, m_A)$.

Figure 1: Timing of the Model

are evaluated by capturing, for example, their business judgment, customer satisfaction rates, or skills as a team worker. The result of the evaluation is used for compensation purposes.

Suppose the principal receive a signal about the agent's performance. For example, she listens to customers praising the employee. She observes the agent at work or the agent's output. She talks to the agent's colleagues to learn about his skills as a team worker. This closely captures a practical evaluation process as "an appraiser would use evidence from direct observation of the employee, or by reports from others, to make judgment about the appraisee's performance." (Porter et al., 2008, p. 149) This signal is subjective and private information of the principal.

Arrow Electronics requires managers to communicate evaluations. Suppose the principal can choose either to tell the agent only the result of the evaluation or to justify her evaluation. Providing justification is costly because it requires the principal to spend additional time and effort on the evaluation.[7] The next section formalizes these notions.

3.2 Setting

Consider a risk-averse agent (he) working for a risk-neutral principal (she). The principal proposes a contract to the agent. The contract specifies payments depending on messages, as described later. After signing such a contract, the agent exerts effort $e \in [0, 1)$, which is unobservable by the principal. Next, the principal privately observes the agent's performance $p \in \mathcal{P} = \{1, 2, \ldots, n\}$ with $n \geq 3$. The performance p is drawn from a distribution $F(p|e) = eF^H(p) + (1 - e)F^L(p)$ depending on the agent's effort e. The distributions $F^H(p)$ and $F^L(p)$ have associated probability measures $f^H(p), f^L(p) > 0$ for all $p \in \mathcal{P}$. In addition, the ratio $f^H(p)/f^L(p)$ strictly increases in p. Therefore, the distribution $F(p|e)$ satisfies the monotone likelihood ratio property, ensuring that higher performance indicates greater work effort.

The principal communicates the evaluation p by sending a message m_P. For this purpose, she has two options: she either provides justification or she does not. If she does not provide

[7] Assume that the agent quits his job at Arrow Electronics afterwards. Turnover rates at Arrow Electronics could reach 20%-25% a year.

justification, her message space is $\mathcal{P}$. If she provides justification, the principal spends costs κ to send a message $m_P \in \{m \in \mathcal{P} | m \geq p\}$.[8] Hence, for good performance, low messages become unavailable. Let $\beta \in \{0, 1\}$ denote the principal's justification decision. For $\beta = 1$, she justifies her evaluation of the agent's work. After observing the principal's choice β and her message m_P, the agent replies with an unverifiable message $m_A \in \{0, 1\}$. Both parties can lie and send any message from the corresponding message spaces. Finally, the contract is performed according to the messages m_P and m_A. The contract specifies the payments made by the principal $W(m_P, m_A)$ and the agent's wage $C(m_P, m_A)$ depending on these two messages. As MacLeod (2003, Proposition 2), Fuchs (2007, Proposition 1) and MacLeod and Tan (2017, Section 2.3) demonstrate, some surplus must be destroyed in this type of model to implement positive effort by the agent. An alternative is to use stochastic contracts, as in Lang (2018). I follow the first approach and allow for $W(m_P, m_A) \geq C(m_P, m_A)$. The principal has no commitment power other than the contract. Figure 1 summarizes the timing.

There is an increasing function $B\colon \mathcal{P} \to [0, \infty)$, such that the principal's benefit is $B(p) - w - \beta\kappa$ if she pays a wage w. The agent's preferences are represented by $u(w) - d(e)$ if he chooses effort e and receives a wage w. I assume $\lim_{w \to 0} u(w) = -\infty$ with derivatives $u' > \epsilon > 0$ and $u'' < 0$. The disutility d of exerting effort is increasing and strictly convex with $d'(0) = 0$ and the limit $\lim_{e \to 1} d(e) = \infty$. Both functions are twice continuously differentiable. The agent receives a reservation utility $\bar{u}$ if he rejects the contract.

3.3 Analysis

Grossman and Hart (1983) show that the model can be solved in two steps. First, for every level of effort e, an optimal contract and its expected costs $\Pi(e)$ for the principal are computed. The second step determines optimal effort levels e by solving

$$\max_{e \in [0,1)} \sum_{p \in \mathcal{P}} B(p) f(p|e) - \Pi(e).$$

Returning to the first step, I determine an optimal contract that implements effort e. Therefore, I consider contracts *feasible* if the agent accepts them, and the contract incentivizes the agent to choose effort e as well as $W(m_P, m_A) \geq C(m_P, m_A)$ for any combination of messages.

Begin with a benchmark that provides a lower bound on the principal's expected costs for any feasible contract.

Lemma 1. *If the evaluation p is observable and contractible, optimal wages $w_e^*(p)$ depend only on the evaluation p. For positive effort $e > 0$, wages $w_e^*(p)$ increase in the evaluation p. There is no justification.*

[8] My analysis does not depend on this particular message space. The analysis is valid for any message space $\mathcal{M}_P$ that is sufficiently rich and shrinks in the evaluation p, i.e., $\mathcal{M}_P(p) \subset \mathcal{M}_P(p')$ for all $p > p'$.

As Holmström (1979) shows, with contractible information, the trade-off between insurance and incentives implies that better evaluations yield higher wages. Therefore, the principal uses upward-sloping wage payments. The monotone likelihood ratio property implies that paying higher wages for better evaluations incentivizes the agent to exert the appropriate work effort.

If the principal's information is subjective and the principal can choose whether to provide justification, messages and justification choices do matter. If there are no justification costs, these additional incentives change optimal contracts but do not change equilibrium wages.

Proposition 1. *If there are no justification costs and $\kappa = 0$, there is an optimal contract in which all evaluations are justified. Equilibrium wages are the same as in Lemma 1. The contract*

$$C(m_P, m_A) = w_e^*(m_P) \quad \textit{and} \quad W(m_P, m_A) = \begin{cases} w_e^*(m_P) & \textit{if } m_A = 1 \\ w_e^*(n) & \textit{else} \end{cases}$$

is optimal.

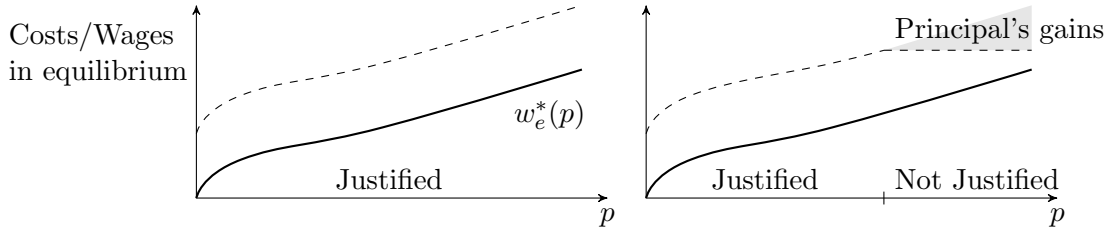
The proposition constructs the optimal contract that is unique in terms of equilibrium utilities. The contract differs from the contract in Lemma 1. In particular, payments depend on whether the agent agrees, i.e., $m_A = 1$. The agent agrees if he observes and reports justification. On the equilibrium path, the agent agrees. Therefore, the equilibrium wage is $w_e^*(p)$ — the same as before. If the agent agrees, the principal's payment equals the agent's wage.

If the principal provides justification, her scope for distorting the evaluation is limited. If the agent receives a poor evaluation, the principal might have distorted the evaluation downwards. A justification, however, guarantees that the principal did not distort the evaluation downwards because low messages become unavailable for good performance. The reporting incentives for the principal are as follows: If she does not provide justification, the agent reports $m_A = 0$, and the principal's payments increase to $w_e^*(n)$. Thus, providing justification is optimal. In addition, reporting $m_P > p$ increases her payments, while reporting $m_P < p$ is impossible. Hence, providing justification and reporting $m_P = p$ is optimal for the principal.

Whenever justification is costly and $\kappa > 0$, however, it is no longer optimal to justify all evaluations. To gain some intuition, suppose, to the contrary, that the principal justifies all evaluations; then, optimal contracts imply wages $w_e^*(p)$ in equilibrium as in Proposition 1. The principal can modify this contract to save on justification costs. The reason is that the agent does not suspect a distorted evaluation by the principal for the highest wages. Therefore, it is not optimal to justify all evaluations.

Lemma 2. *If justification is costly and $\kappa > 0$, justifying all evaluations is not optimal: In an optimal contract, there is a $p \in \mathcal{P}$ with $\beta(p) = 0$.*

The proof shows that the principal's total costs decrease if she refrains from justifying the best evaluations. By paying a high wage that is not justified, she can reduce justification costs.



The left-hand side illustrates the optimal contract if all evaluations are justified. The right-hand side shows a contract in which all but the best evaluations are justified. Thick lines denote the agent's wage in equilibrium, while dashed lines depict the principal's total costs, including justification costs. On the left-hand side, the agent's wage equals $w_e^*(p)$, and the principal's total costs are $w_e^*(p) + \kappa$. On the right-hand side, the agent's wages remain unchanged, but the principal makes third-party payments for good evaluations and pays justification costs only for poor evaluations. For ease of exposition, I draw wages as continuous, although they are discrete.

Figure 2: Idea of the Proof of Lemma 2

These efficiency gains go partly to the principal, as indicated by the gray areas in Figure 2. As demonstrated in the proof, the new contract is feasible. Therefore, it is not optimal to justify all evaluations. It remains to determine which evaluations to justify. To find the optimal justification strategy, however, we need to know more about optimal contracts. It proves convenient to characterize contracts in terms of equilibrium utilities.[9] In a first step, we derive a tighter lower bound for the principal's expected payments.

Lemma 3. *Any solution of Program B that yields implementable equilibrium utilities solves the principal's problem:*

$$\min \sum\nolimits_{p\in\mathcal{P}} (w(p) + \kappa\beta(p)) f(p|e), \tag{B}$$

$$\text{subject to } \sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big) f(p|e) - d(e) \geq \bar{u}, \tag{PC}$$

$$\sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big)\big(f^H(p) - f^L(p)\big) \geq d'(e), \tag{IC}$$

$$\bar{w} \geq w(p) + \kappa\beta(p) \geq (1 - \beta(p))\bar{w} \qquad \forall p \in \mathcal{P} \tag{1}$$

$$w(p) + \kappa\beta(p) \text{ non-decreasing in } p \tag{2}$$

$$w(p) \geq c(p) \qquad \forall p \in \mathcal{P} \tag{3}$$

For effort e, Program B determines equilibrium utilities and justification in an optimal contract. The objective is to minimize expected costs subject to five conditions: The participation constraint PC makes the agent accept the proposed contract. The agent's incentive compatibility IC guarantees that the agent chooses the desired level of effort. In addition, any implementable equilibrium utilities satisfy constraints (1) and (2). If the principal does not provide justification, the agent cannot verify the evaluation by the principal. Therefore, the principal's payments must be constant in the absence of justification. For the same reason, the principal's payment without justification must be higher than payments with justification. Otherwise, the principal would

[9] I call equilibrium utilities *implementable* if there is a feasible contract that implements these equilibrium utilities.

deviate by not justifying the evaluation and paying the lower payment that requires no justification. The contract cannot detect such a deviation. The high pooling wage guarantees that this deviation is unprofitable. Finally, the principal's payments must be higher than the agent's wage. As the next lemma shows, Program B points to threshold rules as optimal communication. Hence, the principal justifies only poor evaluations. Upon receiving a poor evaluation, the agent suspects a distortion by the principal, who alleviates these suspicions by providing justification.

Lemma 4. *In any solution of Program B, there is a* $\delta \in \{0, 1, \dots, n\}$ *with* $\beta(p) = 1$ *if and only if* $p \leq \delta$.

Before turning to formal arguments, consider the following intuition. If optimal communication did not follow a threshold rule as in Lemma 4, there would be evaluations p_L and p_H such that $p_L < p_H$ and the principal justifies p_H but does not provide justification for p_L. Such a communication pattern implies that equilibrium payments decrease in the evaluation p according to Lemma 3. The monotone likelihood ratio property ensures that decreasing equilibrium payments are not optimal. Hence, a threshold rule is optimal. More formally, such a communication pattern implies that costs κ plus equilibrium payments for evaluation p_H equal equilibrium payments for evaluation p_L. Adjust the contract so that the principal does not justify p_H. At the same time, decrease the agent's wage for p_L and increase the wage for p_H such that the agent's expected utility remains constant. Hence, the agent's participation constraint PC is still satisfied. I show that the agent's incentive compatibility IC is slack in this modified contract due to the monotone likelihood ratio property. The proof then shows that this slackness allows a decrease of the principal's costs. Therefore, the initial communication pattern cannot be optimal, and the principal justifies only poor evaluations. Looking a few steps ahead, Proposition 2 shows that such a communication pattern, as depicted in Figure 3, is indeed optimal for the principal. To determine whether it is optimal to justify any evaluations at all, consider the optimal contract when the principal provides no justifications.

Lemma 5. *If the principal does not justify any evaluations, the following contract is optimal:*

$$C(m_P, m_A) = \begin{cases} u^{-1}\left(\bar{u} + d(e) - \frac{(1-f(1|e))d'(e)}{f^L(1)-f^H(1)}\right) & \text{if } m_P = 1 \\ u^{-1}\left(\bar{u} + d(e) + \frac{f(1|e)d'(e)}{f^L(1)-f^H(1)}\right) & \text{else} \end{cases}$$

and $W(m_P, m_A) = u^{-1}\left(\bar{u} + d(e) + \frac{f(1|e)d'(e)}{f^L(1)-f^H(1)}\right)$ *for all* $m_P \in \mathcal{P}$ *and all* $m_A \in \{0, 1\}$. *This contract is unique in terms of equilibrium utilities.*

If the principal reports the worst evaluation $m_P = \min \mathcal{P} = 1$, the agent receives a low wage and the remainder of the principal's payments goes to a third party. For all other evaluations, the principal pays the agent only and pays him a high pooling wage. The next lemma shows that such a contract is only optimal if justification costs are prohibitively high.

Lemma 6. *The principal optimally justifies some evaluations if she wants to implement positive effort $e > 0$ and justification costs κ are at most*

$$\bar{\kappa} = u^{-1}\left(\bar{u} + d(e) + \frac{f(1|e)d'(e)}{f^L(1) - f^H(1)}\right) - u^{-1}\left(\bar{u} + d(e) - \frac{(1 - f(1|e))d'(e)}{f^L(1) - f^H(1)}\right).$$

The principal does not justify any evaluations if justification costs κ are higher than $\bar{\kappa}$.

If justification costs κ are prohibitively high, it is not optimal to use justification and, hence, $\delta = 0$. For moderate justification costs, the principal justifies some evaluations, and the justification threshold is at least 1. Before we can determine optimal contracts and justification thresholds δ for moderate costs, there is a final step missing to simplify Program B.

Lemma 7. *Suppose $\kappa \leq \bar{\kappa}$. The solution of Program C satisfies $w \geq w(p) + \kappa$ and $w(p)$ increasing in p for all $p \in \{1, 2, \ldots, \delta\}$:*

$$\min_{\delta \in \mathcal{P}, w(p), w} \sum_{p=1}^{\delta} (w(p) + \kappa) f(p|e) + (1 - F(\delta|e))w, \tag{C}$$

$$\text{subject to } \sum_{p=1}^{\delta} u\big(w(p)\big) f(p|e) + (1 - F(\delta|e))u(w) - d(e) \geq \bar{u}, \tag{PC}$$

$$\sum_{p=1}^{\delta} u\big(w(p)\big)(f^H(p) - f^L(p)) + u(w) \sum_{p=\delta+1}^{n} (f^H(p) - f^L(p)) \geq d'(e) \tag{IC}$$

The lemma states that the monotone likelihood ratio property and optimal communication together imply constraints (1) and (2) in Program B. The wages in equilibrium reflect only the participation constraint and the incentive compatibility. It is the monotone likelihood ratio property that ensures that $w(p)$ increases in p. The pooling wage is higher than the wages with justification because of the definition of $\bar{\kappa}$ and the communication pattern established in Lemma 4. Combining these results allows the derivation of optimal contracts.

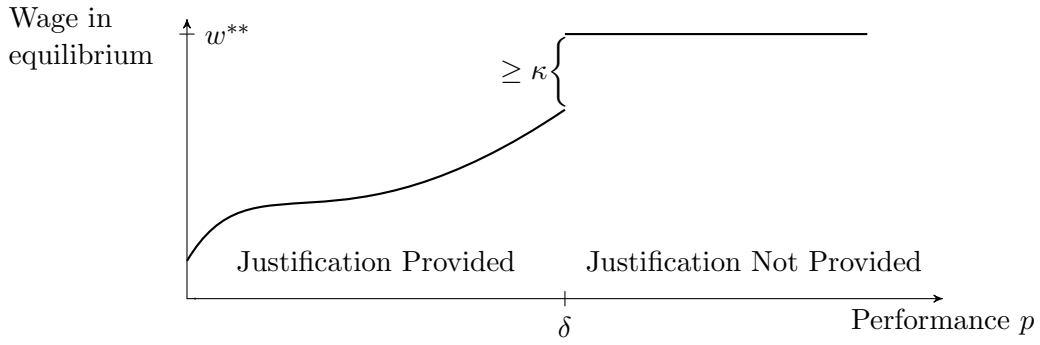
Proposition 2. *Suppose $\kappa \leq \bar{\kappa}$. The following contract is optimal:*

$$C^*(m_P, m_A) = \begin{cases} w^{**} & \text{if } m_P > \delta \\ w^{**}(m_P) & \text{if } m_P \leq \delta \end{cases}$$

$$W^*(m_P, m_A) = \begin{cases} w^{**}(m_P) & \text{if } m_P \leq \delta \text{ and } m_A = 1 \\ w^{**} & \text{else.} \end{cases}$$

Program C in Lemma 7 determines the values of δ, w^{**} and $w^{**}(p)$ for $p \leq \delta$.

The agent's wage depends only on the principal's message m_P. For good evaluations, the wage is constant and equals w^{**}. The principal does not justify these good evaluations, and the agent's wage equals the principal's payments. Varying the agent's wage in these cases means



The optimal justification strategy follows a threshold rule. Poor evaluations are justified, and good evaluations are reported without justification. For ease of exposition, I draw wages as continuous, although they are discrete.

Figure 3: Equilibrium Wages and Communication

reducing the agent's wages. This reduction might ease the agent's incentive compatibility IC but makes it more difficult to satisfy the agent's participation constraint PC. The principal's costs remain unchanged. Instead of varying the agent's wages, however, it is better for the principal to pay a constant wage for good evaluations and to justify poor evaluations, i.e., evaluations with low likelihood ratios. Therefore, the agent's wage equals the principal's payments for good evaluations without justification. For $\kappa > \bar{\kappa}$, there are no justifications at all, and Lemma 5 shows that the agent's wages sometimes differ from the principal's payments.

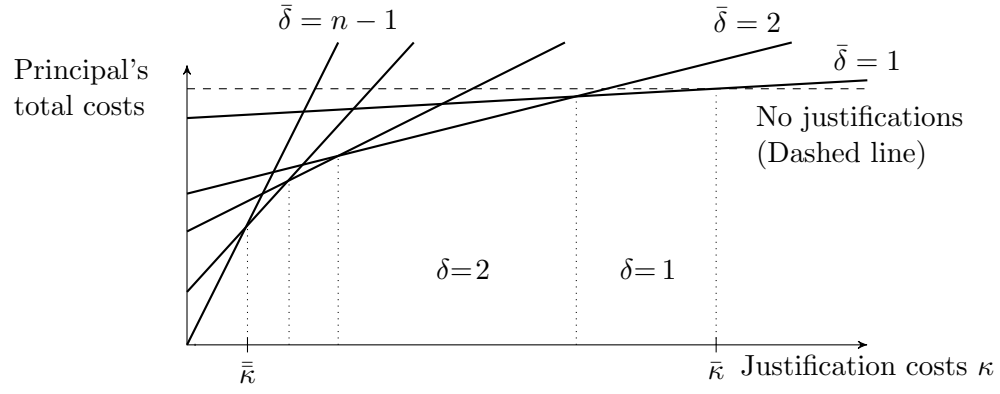
The principal justifies any variation in the agent's wages. She does not justify the highest wage w^{**}. I postpone the discussion of this communication pattern to Corollary 1. If the principal reports an evaluation below δ, her payments depend also on the agent's message. In particular, her payments could differ from the agent's wage. On the equilibrium path, this case never occurs. The agent has the choice of whether to confirm the principal's evaluation. He does so by reporting $m_A = 1$. In equilibrium, he sends such a message whenever he observes a justification by the principal. We can interpret the agent's message as what the agent has observed regarding β.

I provide a feasible contract that implements the equilibrium utilities of Program C. Note that Program C is a simplification of Program B for three reasons: Third-party payments can be avoided on the equilibrium path. Lemma 4 implies that justification follows a threshold rule. Constraint (1) in Program B specifies a constant wage for good evaluations. Moreover, Lemma 7 implies that any solution to Program C satisfies constraints (1) and (2) for $\kappa \leq \bar{\kappa}$. The next section discusses the main features of optimal contracts and strengthens the intuition.

3.4 Properties of Optimal Contracts

Now turn to the properties of optimal contracts for $\kappa \leq \bar{\kappa}$, as determined in Proposition 2. Begin with the communication pattern.

Corollary 1. *The principal optimally justifies evaluations p up to a threshold $\delta \in \{1, 2, \ldots, n-1\}$, while she does not justify evaluations above δ.*



The principal's total costs consist of wage payments and justification costs. The lines depict the principal's total costs for a given justification threshold $\bar{\delta}$. Ex-ante, the principal provides justification with probability $F(\bar{\delta}|e)$. Therefore, her expected justification costs are $F(\bar{\delta}|e)\kappa$, and the slope of the lines is $F(\bar{\delta}|e)$. The pointwise minimum determines the optimal justification threshold δ. The optimal threshold is $\delta = n-1$ for low justification costs $\kappa \leq \bar{\bar{\kappa}}$, and the principal justifies all except the best evaluation. With higher justification costs, the threshold decreases, and the principal justifies fewer evaluations.

Figure 4: Optimal Justification Thresholds

The principal justifies only poor evaluations and low wages; she remains silent on good performance and the best wage because the agent has reason to suspect a distortion only for poor evaluations. Remember that good evaluations imply higher wages and are thus more expensive for the principal. Therefore, the principal never gains by deviating to a better evaluation. Moreover, optimal contracts combine this communication pattern with rewarding evaluations at the top similarly. Thus, they eliminate wage differences that otherwise would call for justification. I describe this optimal communication pattern in the introduction and summarize it in Figure 3. Lemma 4 provides further intuition and, together with Proposition 2, the formal arguments. If justification costs decrease, more evaluations are justified. Recall from Lemma 2 that unless justification is completely costless, however, it is never optimal to justify all evaluations.

Proposition 3. *Suppose $e > 0$. The justification threshold δ and the probability of providing justification decrease in justification costs κ. For sufficiently small costs, the principal justifies all except the best evaluation. Hence, $\delta = n - 1$.*

The threshold δ depends on justification costs κ, as depicted in Figure 4. The lower the justification costs are, the higher the threshold is. Hence, the probability that the principal provides justification increases. Intuitively, the higher the justification costs are, the more it becomes worthwhile to distort the wage payments by pooling more wages at the top to use fewer justifications. The proposition allows for rich comparative statics of the communication pattern to be tested across and within organizations. In particular, there is usually no binary or bang-bang solution for the optimal threshold. The communication pattern could, in principle, be found in time-use data or the length of written evaluations. Alternatively, one could explore differences in justification costs. These differences might arise from principals who are non-native speakers

in multinational corporations, from cultural distance following a takeover or a merger, or from actual distance between principals and agents in companies with several locations.

Next, return to the empirical observations reviewed in the introduction. A large body of literature discusses distortions in wages set by subjective evaluations. The most prominent findings are leniency and centrality. Leniency means that agents receive the highest wages more often than the best performance occurs. Centrality means that variation in performance exceeds variation in wages, particularly at the top. For example, Suvorov and van de Ven (2009, p. 666) state that "compression of performance ratings is well documented." According to Bretz et al. (1992), 60–70% of employees receive an evaluation from the best or second-best category. In addition, Murphy (1993, p. 56) reports that the top 1% of employees at a pharmaceutical company receive a pay increase just 3% higher than the median employee. Taylor and Wherry (1951, p. 39) were the first to find leniency and "a marked distortion ... with considerably poorer discrimination at the top." More recently, Puhani and Yang (2017), Kampkötter and Sliwka (2018), Golman and Bhatia (2012), and Spence and Keeping (2011) also document and confirm leniency and centrality of wages set by subjective evaluations. This literature offers a plethora of approaches and data sources to test leniency and centrality.

Both effects vanish in evaluations for development or feedback instead of wage-setting, as shown, e.g., by Dessler (2008, p. 356), Newman et al. (2017, p. 404), and Jawahar and Stone (1997). This observation is in line with my predictions. The principal requires incentives to report her evaluation for wage-setting truthfully. These incentives cause pooling of the best wages. If the evaluation is for development or feedback, these incentives are unnecessary because the preferences of the principal and the agent with respect to the allocation of training are likely to be better aligned than those with respect to wages. Managers at Merck, for example, reported that "the salary link made discussions on performance improvement difficult." (Murphy, 1993, p. 58) Psychological costs of supervisors giving poor evaluations yield no straightforward explanation of this pattern since those costs should apply to evaluations for all purposes similarly.

In terms of theoretical contributions to this literature, Suvorov and van de Ven (2009) study an intrinsically motivated agent who learns about his productivity from the principal's subjective evaluations. They show that leniency and centrality can be optimal to increase the probability that the agent continues to exert effort. Giebe and Gürtler (2012) show that leniency can be optimal if supervisors are altruistic. For this purpose, they analyze a three-tiered hierarchy. For some parameter values, the principal allows the altruistic supervisor in the optimum to be lenient when evaluating the agent. Indirectly, the papers I discuss in the related literature section under the headings of trust and fairness can also be interpreted as explaining contractual features as the result of altruistic preferences. I do not assume altruistic preferences or intrinsic motivation but instead limit myself to standard preferences. Nonetheless, my optimal contracts exhibit leniency and centrality.

Begin with centrality. Centrality means that, among top performers, variation in performance exceeds variation in wages: $\text{Var}(p|p > \delta) > \text{Var}(\text{equilibrium wages}|p > \delta)$. Alternatively, I can compare the variation of wages at the top derived from my model to the variation of wages at the top for objective and verifiable performance measures, as derived in Lemma 1. Both approaches yield the same result:

Corollary 2. *For $\kappa > \bar{\bar{\kappa}}$, wages set by subjective evaluations exhibit centrality.*

Hence, optimal contracting by a fully rational and unbiased principal results in centrality of wages and wage compression at the top. No biases or non-standard preferences are required to explain centrality in wages.

Now turn to leniency. Leniency means that agents receive the highest wages more often than the best performance occurs: $\Pr(p = n) < \Pr(\text{Equilibrium wages} = \max_{m_P, m_A} C(m_P, m_A))$. Alternatively, I can compare the probability of the highest wages in my model to the probability of the highest wages for objective and verifiable performance measures, as derived in Lemma 1. Both approaches yield the same result:

Corollary 3. *For $\kappa > \bar{\bar{\kappa}}$, wages set by subjective evaluations exhibit leniency.*

The statement of Corollary 3 also remains valid if we extend the definition of leniency and compare the probabilities for the best two performance levels and the highest two wages or for the best three performance levels and the highest three wages, and so on. These results are in line with the aforementioned empirical observations of leniency and centrality: there is less variation in wages set by subjective evaluations than in the underlying performance, in particular at the top. This behavior is *not* the result of a bias but is implied by optimal contracting by a fully rational and unbiased principal, which pools several evaluations and rewards them similarly. Thus, the contract eliminates wage differences that the principal would have to justify. For this reason, I refer to 'leniency' and 'centrality' instead of 'leniency bias' and 'centrality bias'.

Finally, briefly consider leniency and centrality in evaluations directly. To talk about leniency and centrality in evaluations in a meaningful way, it makes sense to think about indirect implementation. In addition, this indirect implementation simplifies optimal contracts. For this purpose, consider the following indirect mechanism: The principal proposes a wage and can provide justification. The agent is then asked whether he accepts or rejects the principal's offer. If the agent accepts the offered wage, the principal pays the wage to the agent. If the agent rejects the offered wage, there is a costly dispute resolution with the costs paid by the principal. These costs can be interpreted as lawyers' fees, mediation costs, or different types of conflicts. Note, however, that it is also possible to implement optimal contracts without such costs and with ex-post budget balance, as shown by Lang (2018). Essentially, the agent can object to the principal's evaluation. This conflict resolution appears realistic because Bretz et al. (1992, p. 332) state that "most organizations report having an informal dispute resolution system (e.g., open door

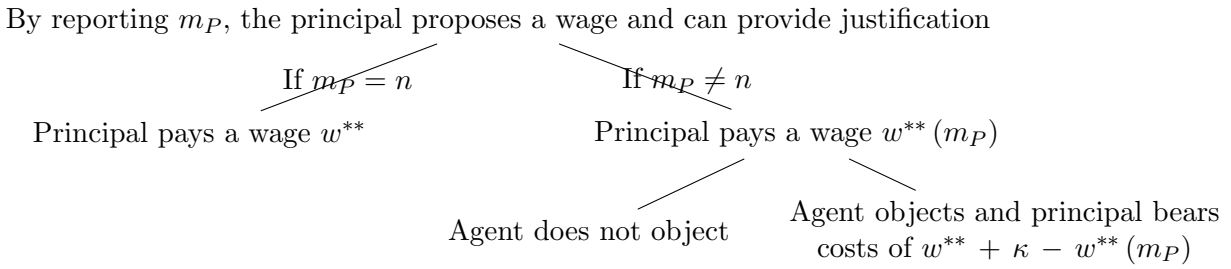


Figure 5: Indirect Implementation of Optimal Contracts

policies) that employees may use to contest the appraisal outcome. About one-quarter report having formalized processes." Figure 5 sketches this implementation. This indirect implementation corresponds to the following contract:

$$C(m_P, m_A) = C^*(m_P, m_A) \quad \text{and} \quad W(m_P, m_A) = \begin{cases} w^{**} & \text{if } m_P = n \\ w^{**}(m_P) & \text{if } m_P \leq \delta \text{ and } m_A = 1 \\ w^{**} + \kappa & \text{else.} \end{cases}$$

It is easy to verify that the contract implements the same incentives and the same utilities for the principal and the agent as optimal contracts in Proposition 2. Indeed, the agent's wages are the same as in Proposition 2. Therefore, the agent chooses the same effort as before. The principal optimally reports $m_P = n$ if the evaluation p is above the threshold δ. Otherwise, she justifies the true evaluation. The agent optimally replies with the message $m_A = 1$ if the principal provides justification. Otherwise, the agent replies with $m_A = 0$. These more realistic contracts also imply leniency and centrality in evaluations. For $\kappa > \bar{\bar{\kappa}}$, evaluations by a fully rational and unbiased principal exhibit centrality, $\text{Var}(p|p > \delta) > \text{Var}(m_P|p > \delta)$, and leniency, $\Pr(p = n) < \Pr(m_P = n)$. Additionally, a stronger notion of leniency is valid here with leniency defined in terms of expectations $\mathbb{E}(m_P) > \mathbb{E}(p)$.

Note, however, that the meaning of messages in messaging games occurs only in equilibrium. Therefore, the optimal contract is not unique in terms of reporting strategies. My optimal contracts can account for leniency and centrality in evaluations, but it is also possible to write down an optimal contract without leniency and centrality in evaluations. Such a contract still exhibits leniency and centrality in wages as shown in the above corollaries.

Next, consider comparative statics. Sometimes, it may be empirically easier to measure differences in information than variation in justification costs. In these cases, costs are a random variable for the observer. Therefore, assume that costs κ are drawn randomly and revealed to the principal and the agent in period -1.[10] Measuring information in moral hazard settings is not trivial. I consider two measures of information here: the worst performance becoming more likely and more precise information for the principal. Begin with the first case. Formally, the

[10] I assume a continuous distribution on $(0, \infty)$ with positive density everywhere.

worst performance becoming more likely means adjusting the distribution $F(p|e)$ in the following way: For an $\epsilon \geq 0$, define $\bar{f}^H(1) = f^H(1) + \epsilon$ and $\bar{f}^L(p) = f^L(p)$ for all $p \in \mathcal{P}$ while reducing the probability measure $\bar{f}^H(p)$ for some better performance levels.[11] For sufficiently small ϵ, $\bar{f}^H(p)$ is a probability measure, and the distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property. Such a change in information yields more contracts with justification and more variation in payments.

Corollary 4. *If the worst performance becomes more likely, i.e., ϵ increases, the principal uses contracts with justification more frequently. In these cases, variation in wages and payments by the principal increases.*

The intuition is that wages increase to compensate for the additional probability of receiving the lowest wage. The concavity of the agent's utilities and the changes in the probabilities imply that the variation in wages increases to incentivize the agent to exert effort. More variation in wages makes providing justification more attractive. Therefore, we observe more contracts with justification.

The second measure of information is the precision of the principal's information in terms of a mean-preserving spread in the likelihood ratios of the agent's performance. Formally, this means adjusting the distribution $F(p|e)$ in the following way: Keep $\bar{f}^H(p) = f^H(p)$ for all $p < n-1$ and $\bar{f}^L(p) = f^L(p)$ for all $p \in \mathcal{P}$ unchanged and adjust $\bar{f}^H(n-1) = f^H(n-1) - \epsilon$ and $\bar{f}^H(n) = f^H(n) + \epsilon$ for an $\epsilon \geq 0$. This adjustment increases the variation in likelihood ratios, improving the principal's information. For sufficiently small ϵ, $\bar{f}^H(p)$ is a valid probability measure and the distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property. More precise information implies more justifications.

Corollary 5. *If the principal's information becomes more precise, i.e., ϵ increases, the principal provides justification more frequently.*

More precise information for the principal reduces her wage costs to incentivize the agent given enough justification. At the same time, the costs to incentivize the agent given little justification remain unchanged because the principal cannot use her improved information without justifications. Therefore, the principal justifies more evaluations for more precise information.

Instead of varying the information, consider the case of the agent's work becoming more demanding. More demanding work means the disutility of exerting effort increases: $\bar{d}(e) = (1+\epsilon)d(e)$ for all $e \in [0, 1)$ and for an $\epsilon > 0$. More demanding tasks yield more contracts with justification and more variation in payments.

Corollary 6. *If the agent's work becomes more demanding, i.e., ϵ increases, the principal uses contracts with justification more frequently. In these cases, variation in wages and payments by the principal increases.*

[11] For example, $\bar{f}^H(p) = f^H(p) - \epsilon/(n-1)$ for all $p > 1$.

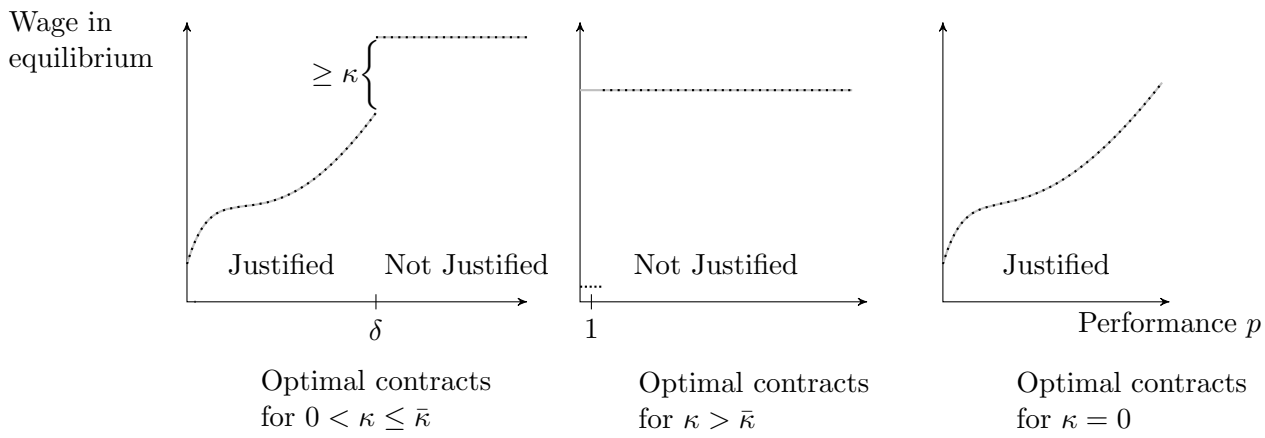
The intuition is that the variation in wages increases to incentivize the agent for the more demanding work. More variation in wages makes providing justifications more attractive. Therefore, we observe more contracts with justification.

4 Conclusions

This paper considers communicating a subjective performance measure in a principal-agent model. The principal can justify her evaluation of the agent's work. Providing justification is costly, does not convey additional information about the agent's effort, and does not have a learning or instructing purpose. Nevertheless, the principal optimally justifies some evaluations if justification is not prohibitively costly. In a well-designed contract, justification guarantees that the principal cannot report poor evaluations for good performance. Therefore, providing justification makes the incentives for the agent credible. The principal justifies only poor evaluations. Upon receiving a good evaluation, the agent is happy to earn a high wage and does not suspect distortion by the principal. Upon receiving a poor evaluation, the agent wants to ensure that the principal evaluated him correctly and did not distort the evaluation downwards to save on wage costs. This communication pattern results in pooling and wage compression at the top, as illustrated in Figure 6. These results fit well with empirical observations, often referred to as leniency bias and centrality bias, as discussed in Corollaries 2 and 3 as well as the introduction. I argue that this pattern of wages is a feature of optimal contracting with unbiased agents and no proof of biased behavior per se.

The principal's justification convinces the agent that the principal has evaluated her appropriately. The expectation of an appropriate evaluation motivates the agent ex-ante to implement the specified work effort. Compare this to a naive contract that does not give the principal an incentive to provide justification. In this naive contract, the principal does not justify the evaluation and always reports the evaluation associated with the lowest wage. Anticipating this behavior, the agent is unmotivated to implement any positive work effort. This partially explains the concern of the management literature to ensure credible feedback provision. In addition, the problem of credible evaluations provides a partial answer to Fuchs (2007, p. 1446), who emphasizes the importance of exploring "possible reasons for the existence of communication" between agent and principal. I show that credibility problems cause communication.

Figure 6 shows the relation to the literature. The most interesting case is for moderate justification costs. Then, optimal contracts and the equilibrium wage pattern are new results and contribute to the literature. In extreme cases of no justification costs and prohibitively large justification costs, equilibrium wages are similar to the previous literature, such as Holmström (1979), Grossman and Hart (1983), MacLeod (2003) or Fuchs (2007), although optimal contracts differ due to my justifications. The middle graph depicts optimal equilibrium wages for prohibitively high justification costs. Unsurprisingly, the principal provides no justifications in this



The dotted lines denote the agent's wage in equilibrium, while the gray lines depict the principal's payments in equilibrium. For ease of exposition, I draw wages as continuous, although they are discrete.

Figure 6: Comparison of optimal contracts

case and her payments are constant in the evaluation. The graph on the right-hand side depicts optimal equilibrium wages for zero justification costs. Unsurprisingly, the principal then provides justifications for all evaluations, and the equilibrium wages and payments increase in the evaluation.

The results of this paper are important for the design of incentive systems. First, the systems must ensure credible provision of appropriate feedback by institutionalizing the feedback process or using multi-source feedback. Second, pooling at the top could cause substantial costs for an incentive scheme if a large proportion of employees receives positive evaluations. Bernardin and Orban (1990, p. 199) provide the example of the Small Business Administration and NASA introducing a bonus scheme based on subjective evaluations. After more than 50% of eligible employees should receive a bonus, Congress responded with the rule that no more than 25% of employees shall receive a bonus.

A Appendix

Lemma 1 characterizes the optimal contract if the principal's information p is contractible. This contract is unique in terms of equilibrium utilities.

Proof of Lemma 1: Contractible information means $m_P = p$. To fit contractible information into my setting, define contract $\mathcal{W}$ by $W(m_P, m_A) = C(m_P, m_A) = w_e(m_P)$ for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$ with a function w_e: $\mathcal{P} \to \mathbb{R}$. Contract $\mathcal{W}$ implies $\beta(p) = 0$ for all $p \in \mathcal{P}$. Then, Program A below determines how to optimally implement effort e by choosing wages $w_e^*(\cdot)$. The objective is to minimize expected costs subject to two conditions. The participation constraint PC makes the agent accept the proposed contract. The agent's incentive compatibility IC guarantees that the agent chooses the desired level of effort.

$$\min_{w_e^*(\cdot)} \sum\nolimits_{p\in\mathcal{P}} w_e^*(p) f(p|e) \tag{A}$$

$$\text{subject to } \sum\nolimits_{p\in\mathcal{P}} u(w_e^*(p)) f(p|e) - d(e) \geq \bar{u} \tag{PC}$$

$$e \in \arg\max \sum\nolimits_{p\in\mathcal{P}} u(w_e^*(p)) f(p|e) - d(e) \tag{IC}$$

To implement no effort, $e = 0$, set $w_0^*(p) = u^{-1}(\bar{u} + d(0))$ for all $p \in \mathcal{P}$. If the principal wants a positive effort, $e > 0$, the agent's incentive compatibility matters. The first-order approach is valid here because $F(p|e)$ is a linear combination of distribution functions. This implies that the convex distribution function condition is satisfied. According to Grossman and Hart (1983) and Rogerson (1985), the convex distribution function condition in combination with the convexity of $d(\cdot)$ and the monotone likelihood ratio property guarantees that the first-order approach is valid. Therefore, the agent's incentive compatibility is equivalent to

$$\sum\nolimits_{p\in\mathcal{P}} u(w(p))(f^H(p) - f^L(p)) \geq d'(e). \tag{IC}$$

Denote by $\mathcal{P}_c = \{p \in \mathcal{P} | f^H(p) - f^L(p) \geq 0\}$. Note that the constraint set is nonempty. Take, for example, the contract

$$w(p) = \begin{cases} w_1 & \text{if } p \in \mathcal{P}_c \\ w_2 & \text{else} \end{cases}$$

with w_1 and w_2 determined below. The contract satisfies the agent's incentive compatibility IC if

$$\begin{aligned} d'(e) &= \sum\nolimits_{p\in\mathcal{P}_c} u(w_1)(f^H(p) - f^L(p)) + \sum\nolimits_{p\in(\mathcal{P}\setminus\mathcal{P}_c)} u(w_2)(f^H(p) - f^L(p)) = \\ &= (u(w_1) - u(w_2)) \sum\nolimits_{p\in\mathcal{P}_c} (f^H(p) - f^L(p)) \end{aligned}$$

because $\sum_{p\in\mathcal{P}} f^H(p) - f^L(p) = 0$ and, hence, $\sum_{p\in\mathcal{P}_c} f^H(p) - f^L(p) = -\sum_{p\in(\mathcal{P}\setminus\mathcal{P}_c)} f^H(p) - f^L(p)$. According to the definition of $\mathcal{P}_c$ and the monotone likelihood ratio property, the last sum is positive. Therefore, the equation uniquely determines $u(w_1) - u(w_2)$. The contract satisfies the

participation constraint PC if

$$d(e) + \bar{u} = u(w_1) \sum\nolimits_{p \in \mathcal{P}_c} f(p|e) + u(w_2) \sum\nolimits_{p \in (\mathcal{P} \setminus \mathcal{P}_c)} f(p|e) =$$
$$= u(w_2) + (u(w_1) - u(w_2)) \sum\nolimits_{p \in \mathcal{P}_c} f(p|e)$$

because $\sum_{p \in \mathcal{P}} f(p|e) = 1$ and, hence, $\sum_{p \in \mathcal{P}_c} f(p|e) = 1 - \sum_{p \in (\mathcal{P} \setminus \mathcal{P}_c)} f(p|e)$. Plugging in the above solution for $u(w_1) - u(w_2)$ uniquely determines w_2. Therefore, the constraint set of Program A is nonempty. Moreover, the costs of an optimal contract are lower than $\max\{w_1, w_2\} < \infty$. According to Grossman and Hart (1983), the strict concavity of $u(\cdot)$ ensures that the solution to Program A is unique. Denote the solution to Program A by $w_e^*(p)$.

Optimization with the Lagrange multipliers of the participation constraint ν_1 and the incentive compatibility ν_2 determines the optimal contract as

$$f(p|e) - \nu_1 u'(w_e^*(p)) f(p|e) - \nu_2 u'(w_e^*(p))(f^H(p) - f^L(p)) = 0,$$
$$\frac{1}{u'(w_e^*(p))} = \nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} = \nu_1 + \nu_2 \frac{\frac{f^H(p)}{f^L(p)} - 1}{e\frac{f^H(p)}{f^L(p)} + 1 - e} \tag{4}$$

The Lagrange multiplier ν_2 is positive for the following reason: If $\nu_2 = 0$, then Eq. (4) implies that $w(p)$ is constant in p, violating the incentive compatibility IC. Hence, $\nu_2 > 0$. Since the fraction $\frac{l-1}{el+1-e}$ increases in l, the right-hand side of Eq. (4) increases in $p \in \mathcal{P}$ due to the monotone likelihood ratio property. Therefore, the concavity of $u(\cdot)$ implies that $w_e^*(p)$ increases in $p \in \mathcal{P}$. □

Proof of Proposition 1: The easiest way to study the principal's problem is in terms of equilibrium utilities and certainty equivalents, respectively. Define expected payments by the principal given equilibrium strategies and evaluation p by $w(p)$. Similarly, define the certainty equivalent of the agent's wages given equilibrium strategies and evaluation p by $c(p)$. The agent is unaware of p, thus $c(p)$ is a purely theoretical concept to analyze the contract. Any feasible contract must satisfy the following three conditions:

1.) The agent accepts the contract if

$$\sum\nolimits_{p \in \mathcal{P}} u\big(c(p)\big) f(p|e) - d(e) \geq \bar{u}. \tag{5}$$

2.) The principal's payment must be at least the agent's wage, and the agent is risk averse. Therefore,

$$w(p) \geq c(p) \qquad \forall p \in \mathcal{P}. \tag{6}$$

3.) The agent implements effort e if

$$e \in \arg\max \left(\sum\nolimits_{p \in \mathcal{P}} u\big(c(p)\big) f(p|e) - d(e) \right).$$

By the same arguments as in Lemma 1, the first-order approach is valid here, and focusing

on the equilibrium utilities $c(p)$ is without loss of generality. Therefore, the agent implements effort e if

$$\sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big)\big(f^H(p)-f^L(p)\big)\geq d'(e). \tag{7}$$

Implementing these equilibrium utilities might require additional constraints, which I neglect for the moment. These additional constraints cannot make the contract cheaper for the principal. Therefore, the principal must pay at least $\min_{w(\cdot)}\sum_{p\in\mathcal{P}} w(p)f(p|e)$, subject to constraints (5), (7), (6). It is easy to see that in any solution of this program, $w(p)=c(p)$ for all $p\in\mathcal{P}$. Hence, the program is equivalent to program A in Lemma 1. Consequently, a lower bound for the principal's expected costs is given by $\sum_{p\in\mathcal{P}} w_e^*(p)f(p|e)$, with $w_e^*(p)$ defined in Lemma 1.

Consider the following contract: $C(m_P,m_A)=w_e^*(m_P)$ and

$$W(m_P,m_A)=\begin{cases} w_e^*\left(m_P\right) & \text{if } m_A=1\\ w_e^*\left(n\right) & \text{else}\end{cases}$$

for all $m_P\in\mathcal{P}$ and all $m_A\in\{0,1\}$. The contract induces the agent to report $m_A=\beta$ because the agent's wage is independent of her message m_A. Hence, we can interpret the agent's message as what the agent has observed regarding β. Whenever the principal does not provide justification, the agent sends the message $m_A=0$, and the principal's payments increase to $w_e^*\left(n\right)$. If the principal deviates to a message $m_P>p$ with justification, her payments increase to $w_e^*(m_P)$. Lemma 1 ensures that these payments are bigger than $w_e^*(p)$ because $w_e^*(\cdot)$ is increasing. Hence, the contract induces the principal to report $m_P=p$ and $\beta(p)=1$ for all $p\in\mathcal{P}$. It is easy to see that the principal's payments are always higher than the agent's wage. The contract proposed above implements the equilibrium wage $w_e^*(p)$. More importantly, the contract is optimal because it is feasible and attains the lower bound on the principal's costs. Since it is impossible to incentivize the agent with a cheaper contract, this contract is optimal for $\kappa=0$. □

Proof of Lemma 2: Suppose, to the contrary, that the principal justifies all evaluations: $\beta(p)=1$ for all $p\in\mathcal{P}$. Then, following the arguments in Proposition 1, the principal's expected costs are at least $\kappa+\mathbb{E}(w_e^*(p))$ with $w_e^*(\cdot)$ defined in Lemma 1. Nonetheless, I am going to show that partial justification implements effort e more cheaply.

Consider contract $\mathcal{W}$ with $C(m_P,m_A)=w_e^*(m_P)$ and

$$W(m_P,m_A)=\begin{cases} w_e^*(m_P) & \text{if } m_P\neq n \text{ and } m_A=1\\ \max\{w_e^*(n-1)+\kappa, w_e^*(n)\} & \text{if } m_P=n\\ w_e^*(n)+\kappa & \text{else}\end{cases}$$

It is easy to check that the principal's payments are always higher than the agent's wage.

The principal justifies all evaluations except the highest ones: $\beta(n)=0$ and $\beta(p)=1$ for all $p<n$. If the principal's information indicates excellent performance, $p=n$, justification would increase her costs by κ because the wage costs remain unchanged. As in Proposition 1, the contract induces the agent to report $m_A=\beta$ and the principal to report $\beta=\beta(p)$ and $m_P=p$. Hence, the agent's equilibrium wage equals $w_e^*(p)$. Therefore, this contract is cheaper

than $\kappa+\mathbb{E}(w_e^*(p))$ and implements effort e by the definition of $w_e^*(\cdot)$. In particular, the principal's expected costs decrease by

$$\begin{aligned} &f(n|e)(w_e^*(n)+\kappa-\max\{w_e^*(n-1)+\kappa, w_e^*(n)\}) = \\ =&f(n|e)\min\{w_e^*(n)-w_e^*(n-1),\kappa\} > 0. \end{aligned}$$

This shows that the principal should not justify all evaluations for positive costs κ. □

Proof of Lemma 3: As in Proposition 1, we study the principal's program in terms of equilibrium utilities. Remember that expected payments for the principal given equilibrium strategies and evaluation p are denoted by $w(p)$. Similarly, the corresponding certainty equivalents of the agent's wages are $c(p)$. The agent is unaware of p, thus $c(p)$ is a purely theoretical concept to analyze the contract. Finally, denote the principal's equilibrium justification strategy by $\beta(p)$. Now turn to the additional constraints, which I neglected in Proposition 1.

For $\beta(p)=0$, the agent cannot check whether the principal reports the performance p truthfully. Hence, optimal equilibrium play requires that equilibrium payments are constant in the principal's message as I prove here: Suppose, to the contrary, that there are $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1)=\beta(p_2)=0$ and $w(p_1)\neq w(p_2)$. Without loss of generality, assume $w(p_1)>w(p_2)$. If p_1 occurs, the principal could deviate from the equilibrium strategies and use the reporting strategy as if p_2 had occurred. The agent is unaware of this deviation and, in addition, p_2 and the resulting message are consistent with effort e, i.e., the probabilities of p_2 conditionally on e are strictly positive. Hence, the agent does not detect the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A that cannot depend on p because the performance p is the principal's private information. Therefore, the principal expects a wage $w(p_2)$ following this deviation. By assumption, $w(p_1)>w(p_2)$, yielding a contradiction to the optimality of equilibrium play. Consequently, $w(p_1)=w(p_2)$ for all $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1)=\beta(p_2)=0$.

Similarly, optimal equilibrium play requires that the equilibrium costs of the principal are lower with justification than without justification, as I prove here: Suppose, to the contrary, that there are $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1)=0$ and $\beta(p_2)=1$ and $w(p_1)<w(p_2)+\kappa$. If p_2 occurs, the principal could deviate from the equilibrium strategies and use the reporting strategy with $\beta=0$ and m_P as if p_1 had occurred. The agent is unaware of this deviation and, in addition, p_1 and the resulting message are consistent with effort e, i.e., the probabilities of p_1 conditionally on e are strictly positive. Hence, the agent does not detect the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A that cannot depend on p because the performance p is the principal's private information. Therefore, the principal expects costs of $w(p_1)$ following this deviation. By assumption, $w(p_1)<w(p_2)+\kappa$, yielding a contradiction to the optimality of equilibrium play. Consequently, $w(p_1)\geq w(p_2)+\kappa$ for all $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1)=0$ and $\beta(p_2)=1$. Together, optimal equilibrium play implies

$$\bar{w} \geq w(p)+\kappa\beta(p) \geq (1-\beta(p))\bar{w} \qquad \forall p \in \mathcal{P} \tag{8}$$

for a constant $\bar{w}$. Next, optimal equilibrium play requires that the principal's equilibrium costs are non-decreasing in p. Suppose, to the contrary, that there are $p_L, p_H \in \mathcal{P}$ with $w(p_L) + \kappa\beta(p_L) > w(p_H) + \kappa\beta(p_H)$ and $p_L < p_H$. If p_L occurs, the principal could deviate from the equilibrium strategies and use the reporting strategy as if p_H had occured. The agent is unaware of this deviation and the resulting message is consistent with effort e, i.e., the probabilities of p_H conditionally on e are strictly positive. Hence, the agent cannot detect the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A. Therefore, the principal expects costs of $w(p_H) + \kappa\beta(p_H)$ following this deviation. By assumption, $w(p_L)+\kappa\beta(p_L) > w(p_H)+\kappa\beta(p_H)$, yielding a contradiction to the optimality of equilibrium play. Consequently,

$$w(p) + \kappa\beta(p) \text{ is non-decreasing in } p. \tag{9}$$

If these equilibrium utilities are implementable with a feasible contract, they solve the principal's problem. The principal is minimizing her expected costs:

$$\min_{w(p),\bar{w},c(p),\beta(p)} \sum\nolimits_{p\in\mathcal{P}} (w(p) + \kappa\beta(p)) f(p|e)$$

subject to constraints (5), (6), (7), (8), and (9). This program equals Program B. □

Proof of Lemma 4: To show that only poor evaluations are justified, assume to the contrary that there is a solution $\mathcal{W}^{**}$ to Program B with $p_L, p_H \in \mathcal{P}$, such that $p_L < p_H$, $\beta^{**}(p_L) = 0$ and $\beta^{**}(p_H) = 1$. Constraints (1) and (2) imply that $w(p_H) + \kappa = w(p_L)$.

The next step constructs a $\mathcal{W}'$ that implements effort e cheaper than $\mathcal{W}^{**}$. Modify the solution $\mathcal{W}^{**}$ in the following way to get $\mathcal{W}'$: Set $\beta'(p_H) = 0$ and $w'(p_H) = c'(p_H) = w(p_L)$ and $c'(p_L) = \tilde{c}$ with $\tilde{c}$ determined by

$$u(\tilde{c}) = u(c(p_L)) - (u(w(p_L)) - u(c(p_H)))\frac{f(p_H|e)}{f(p_L|e)}.$$

Otherwise, $\mathcal{W}'$ equals $\mathcal{W}^{**}$. The agent's certainty equivalent for an evaluation p_L decreases from $c(p_L)$ to $\tilde{c}$ in $\mathcal{W}'$, while the certainty equivalent for an evaluation p_H increases from $c(p_H)$ to $w(p_L)$, as $c(p_H) \leq w(p_H) < w(p_L)$. The principal's expected costs remain unchanged. In $\mathcal{W}'$, the principal does not justify p_L and p_H. Hence, $\beta'(p_L) = \beta'(p_H) = 0$. The definition of $\tilde{c}$ implies that $\tilde{c} < c(p_L) \leq w(p_L)$ and that the change in the agent's expected utilities ($\mathbb{E}[u(\mathcal{W}') - u(\mathcal{W}^{**})]$ in sloppy notation) equals

$$\begin{aligned}&\sum\nolimits_{i\in\{L,H\}} (u(c'(p_i)) - u(c(p_i)))f(p_i|e) = \\ =&(u(\tilde{c}) - u(c(p_L)))f(p_L|e) + (u(w(p_L)) - u(c(p_H)))f(p_H|e) = 0.\end{aligned} \tag{10}$$

Hence, $\mathcal{W}'$ satisfies the participation constraint PC in Program B because the agent's expected utilities remain unchanged. Additionally, the left-hand side of the incentive compatibility IC is now larger than the marginal cost of effort, $d'(e)$ because the left-hand side of the incentive compatibility IC changes by

$$\sum_{p\in\mathcal{P}}(f^H(p)-f^L(p))(u(c'(p))-u(c(p)))=\sum_{i\in\{L,H\}}(f^H(p_i)-f^L(p_i))(u(c'(p_i))-u(c(p_i)))=$$
$$=(f^H(p_L)-f^L(p_L))(u(\tilde{c})-u(c(p_L)))+(f^H(p_H)-f^L(p_H))(u(w(p_L))-u(c(p_H)))=$$
$$=\frac{f^H(p_L)-f^L(p_L)}{f(p_L|e)}f(p_L|e)(u(\tilde{c})-u(c(p_L)))+\frac{f^H(p_H)-f^L(p_H)}{f(p_H|e)}f(p_H|e)(u(w(p_L))-u(c(p_H)))>$$
$$>\frac{f^H(p_L)-f^L(p_L)}{f(p_L|e)}\left(f(p_L|e)(u(\tilde{c})-u(c(p_L)))+f(p_H|e)(u(w(p_L))-u(c(p_H)))\right)=0.$$

The last equality follows from Eq. (10). The main inequality follows from the monotone likelihood ratio property, which ensures that

$$\frac{f^H(p)-f^L(p)}{f(p|e)}=\frac{\frac{f^H(p)}{f^L(p)}-1}{e\frac{f^H(p)}{f^L(p)}+1-e}\text{ increases in }p\in\mathcal{P}.$$

$\mathcal{W}'$ satisfies constraint (2) because the principal's costs remain unchanged: $w'(p)+\kappa\beta'(p)=w(p)+\kappa\beta(p)$ for all $p\in\mathcal{P}$. In addition, $\mathcal{W}'$ also satisfies constraints (1) and (3).

Apply this procedure repeatedly until only poor evaluations are justified and there are no more $p_L,p_H\in\mathcal{P}$ with the required properties. This procedure transforms solution $\mathcal{W}^{**}$ into $\mathcal{W}^m$ with the communication choice $\beta^m(p)$. Finally, I show how to make $\mathcal{W}^m$ cheaper for the principal. There are two cases to consider: 1.) $\beta^m(1)=1$; 2.) $\beta^m(1)=0$. Begin with the first case and $\beta^m(1)=1$. The above steps ensure that there is a $p_3\in\mathcal{P}$ with $c^m(p_3)<w^m(p_3)$ and $\beta^m(p_3)=0$. Hence, $p_3\neq 1$. Increase $c^m(p_3)$ by a small $\epsilon>0$ to $c^m(p_3)+\epsilon$ and reduce $w^m(1)$ and $c^m(1)$ to $\tilde{c}^\epsilon$ with $\tilde{c}^\epsilon$ determined by

$$u(\tilde{c}^\epsilon)=u(c^m(1))-(u(c^m(p_3)+\epsilon)-u(c^m(p_3)))\frac{f(p_3|e)}{f(1|e)}.$$

This modification does not affect the agent's participation constraint. I showed above that the constraint IC was slack. Hence, choosing ϵ that is sufficiently small ensures that the modified contract satisfies constraints IC and (3) due to $c^m(p_3)+\epsilon\leq w^m(p_3)$. It is easy to see that the modified contract satisfies constraints (1) and (2) because reducing $w^m(1)$ never violates constraint (2). Moreover, $\beta^m(1)=1$ and reducing $w^m(1)$ does not violate constraint (1). The modified $\mathcal{W}^m$ implies lower expected costs for the principal compared to contract $\mathcal{W}^{**}$. In addition, the modified $\mathcal{W}^m$ satisfies all the constraints of Program B. Therefore, the modified $\mathcal{W}^m$ contradicts the optimality of the solution $\mathcal{W}^{**}$.

Next, turn to the second case with $\beta^m(1)=0$. Note that $\beta^m(1)=0$ implies $\beta^m(p)=0$ for all $p\in\mathcal{P}$. Otherwise, there were $p_L,p_H\in\mathcal{P}$ with the required properties. Again, the above steps ensure that there is a $p_3\in\mathcal{P}$ with $c^m(p_3)<w^m(p_3)$. Increase $c^m(p_3)$ by a small $\epsilon>0$ to $c^m(p_3)+\epsilon$ and reduce the remaining $c^m(p)$ and all $w^m(p)$ by $\hat{\epsilon}^\epsilon$ with $\hat{\epsilon}^\epsilon$ determined by

$$f(p_3|e)(u(c^m(p_3)+\epsilon)-u(c^m(p_3)))+\sum_{p\in\mathcal{P}\setminus\{p_3\}}f(p|e)(u(c^m(p)-\hat{\epsilon}^\epsilon)-u(c^m(p)))=0.$$

This modification does not affect the agent's participation constraint. I showed above that the constraint IC was slack. Hence, choosing ϵ that is sufficiently small ensures that the modified

contract satisfies constraints IC and (3) due to $c^m(p_3)+\epsilon \leq w^m(p_3)-\hat{\epsilon}^\epsilon$. It is easy to see that the modified $\mathcal{W}^m$ satisfies constraints (1) and (2) because reducing all $w^m(p)$ equally never violates constraints (1) and (2). The modified $\mathcal{W}^m$ implies lower expected costs for the principal compared to solution $\mathcal{W}^{**}$. In addition, the modified $\mathcal{W}^m$ satisfies all the constraints of Program B. Therefore, the modified $\mathcal{W}^m$ contradicts the optimality of the solution $\mathcal{W}^{**}$. Consequently, justification optimally follows a threshold rule, and there is a $\delta \in \{0,1,\ldots,n\}$ with $\beta(p)=1$ if and only if $p \leq \delta$. □

Proof of Lemma 5: Assume $\beta(p)=0$ for all $p \in \mathcal{P}$. Then, constraint (1) in Program B of Lemma 3 implies $w(p)=\bar{w}$ for all $p \in \mathcal{P}$. Hence, Program B simplifies to

$$\begin{aligned}
&\min_{c(p),\bar{w}} \bar{w},\\
\text{subject to} &\sum\nolimits_{p\in\mathcal{P}} u(c(p))f(p|e) - d(e) \geq \bar{u}, && \text{(PC)}\\
&\sum\nolimits_{p\in\mathcal{P}} u(c(p))(f^H(p)-f^L(p)) \geq d'(e), && \text{(IC)}\\
&f(p|e)(\bar{w}-c(p)) \geq 0 \qquad \forall p \in \mathcal{P}. && (11)
\end{aligned}$$

Define ν_1, ν_2 and $\chi(p)$ as the Lagrange multipliers of the conditions PC, IC and (11), respectively. If $c(p)=\bar{w}$ for all $p \in \mathcal{P}$, the contract violates the incentive compatibility IC. Therefore, there is an evaluation $p^* \in \mathcal{P}$ with third-party payments, i.e., $c(p^*)<\bar{w}$. Then, the complementary slackness condition yields $\chi(p^*)=0$. Optimization of the Lagrangian with respect to $c(p^*)$ results in

$$\begin{aligned}
-\nu_1 u'(c(p^*))f(p^*|e) - \nu_2 u'(c(p^*))(f^H(p^*)-f^L(p^*)) &= 0\\
\nu_1 + \nu_2 \frac{f^H(p^*)-f^L(p^*)}{f(p^*|e)} &= 0. \qquad (12)
\end{aligned}$$

The monotone likelihood ratio property ensures that $\frac{f^H(p)-f^L(p)}{f(p|e)}$ strictly increases in $p \in \mathcal{P}$. In addition, ν_2 is positive in an optimum because the solution to the above program without the incentive compatibility IC is $\bar{w}=c(p)=u^{-1}(\bar{u}+d(e))$ for all $p \in \mathcal{P}$, and this solution violates constraint IC. Therefore, equation (12) can hold for at most one $p^* \in \mathcal{P}$. Hence, $c(p)=\bar{w}$ and $\chi(p) \geq 0$ for all $p \in \mathcal{P} \setminus \{p^*\}$.

Assume, to the contrary, that $p^* \neq 1$. Optimization of the Lagrangian with respect to $c(1)$ results in

$$\begin{aligned}
-\nu_1 u'(c(1))f(1|e) - \nu_2 u'(c(1))(f^H(1)-f^L(1)) + \chi(1)f(1|e) &= 0\\
\nu_1 + \nu_2 \frac{f^H(1)-f^L(1)}{f(1|e)} &= \frac{1}{u'(c(1))}\chi(1).
\end{aligned}$$

Equation (12), $\nu_2>0$, $1<p^*$ and the monotone likelihood ratio property imply that the left-hand side of the last equation is negative. The right-hand side is non-negative because constraint (11) is binding for 1 and $\chi(1) \geq 0$ as $p^* \neq 1$. This contradiction proves that $p^*=1$. Therefore, $c(p)=\bar{w}$ and $\chi(p) \geq 0$ for all $p \in \{2,3,\ldots,n\}$.

Plugging these results into constraints PC and IC yields:

$$u(w)(1-f(1|e)) + u(c(1))f(1|e) - d(e) = \bar{u},$$

$$(u(c(1)) - u(w))(f^H(1) - f^L(1)) \geq d'(e)$$

because $0 = \sum_{p\in\mathcal{P}} f^H(p) - f^L(p) = f^H(1) - f^L(1) + \sum_{p=2}^{n} f^H(p) - f^L(p)$. Solving the first equation for $u(c(1))$ gives

$$u(c(1)) = \frac{\bar{u} + d(e) - u(w)(1 - f(1|e))}{f(1|e)}.$$

Inserting this value for $u(c(1))$ into the second inequality leads to

$$(\bar{u} + d(e) - u(w))\,(f^H(1) - f^L(1)) \geq f(1|e)d'(e)$$

and finally results in

$$u(w) = \bar{u} + d(e) + \frac{f(1|e)}{f^L(1) - f^H(1)}d'(e) \qquad \text{and}$$
$$u(c(1)) = \bar{u} + d(e) - \frac{1 - f(1|e)}{f^L(1) - f^H(1)}d'(e).$$

It remains to show that these equilibrium utilities are implementable with a feasible contract. Consider the contract $W(m_P, m_A) = w$ and

$$C(m_P, m_A) = \begin{cases} c(1) & \text{if } m_P = 1 \\ w & \text{else} \end{cases}$$

for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$. Since the principal's payments are independent of her message, she provides no justification, and reporting $m_P = p$ is optimal for her. Hence, the contract implies the required equilibrium utilities. It remains to verify that this contract is feasible. By definition of w and $c(1)$ as well as the principal's reporting strategy, the agent accepts the contract and chooses effort e. Finally, it is straightforward to check that the principal's payments $W(m_P, m_A)$ are always at least as high as the agent's wage $C(m_P, m_A)$. Consequently, if $e > 0$ and $\beta(p) = 0$ for all $p \in \mathcal{P}$, the contract is optimal for the principal because it is feasible and attains the lower bound of Program B according to Lemma 3. □

Proof of Lemma 6: The proof proceeds in two steps. First, I show that using justification is not optimal if the costs κ are above the threshold in the lemma. Second, I show that for lower costs κ and $e > 0$, there is a feasible contract with $\beta(p) = 1$ for a $p \in \mathcal{P}$ and lower costs for the principal than any contract without justification.

Begin with the first step. If $e = 0$, the optimal contract is $W(m_P, m_A) = C(m_P, c_A) = u^{-1}(\bar{u} + d(0))$ for all m_P and m_A. Hence, it is not optimal to use justification for $\kappa > 0$. Therefore, restrict attention to $e > 0$. If the principal does not provide justification, her costs are

$$w = u^{-1}\left(\bar{u} + d(e) + \frac{f(1|e)}{f^L(1) - f^H(1)}d'(e)\right)$$

according to Lemma 5. Suppose $\kappa > \bar{\kappa}$. Assume to the contrary that there is an optimal contract $\mathcal{W}'$ with some justification, i.e., there is a $\tilde{p} \in \mathcal{P}$ with $\beta'(\tilde{p}) = 1$. Again, denote the principal's expected payments in contract $\mathcal{W}'$ given equilibrium strategies and evaluation p by $w'(p)$ and the agent's certainty equivalents of the wages given equilibrium strategies and evaluation p by

$c'(p)$. Lemma 4 shows that $\beta'(\tilde{p}) = 1$ implies $\beta'(1) = 1$. If $p = 1$, the principal's costs are at least $\kappa + c'(1)$. Lemma 3 proves that $w'(p) + \kappa\beta'(p)$ is non-decreasing in p. Therefore, the principal's costs are at least $\kappa + c'(1)$ for any $p \in \mathcal{P}$. Hence, the costs of contract $\mathcal{W}'$ are at least $\kappa + c'(1) > \bar{\kappa} + c'(1) = w - c(1) + c'(1)$ because $\kappa > \bar{\kappa} = w - c(1)$ with $c(1)$ determined in Lemma 5. If $c(1) \leq c'(1)$, these costs are above w. Thus, contract $\mathcal{W}'$ is more expensive than contract $\mathcal{W}$ in Lemma 5. Consequently, restrict attention to $c(1) > c'(1)$. In contract $\mathcal{W}$, the agent's participation constraint PC was binding. Therefore, the agent's participation constraint PC requires the following for contract $\mathcal{W}'$:

$$\begin{aligned}(1 - f(1|e))u(\mathbb{E}(c'(p)|p \neq 1)) + f(1|e)u(c'(1)) \geq \\ \geq \bar{u} + d(e) = (1 - f(1|e))u(w) + f(1|e)u(c(1)).\end{aligned}$$

Rearranging yields

$$u(\mathbb{E}(c'(p)|p \neq 1)) \geq u(w) + \frac{f(1|e)}{1 - f(1|e)}(u(c(1)) - u(c'(1))) > u(w). \tag{13}$$

Hence, $c'(1) < c(1) < w < \mathbb{E}(c'(p)|p \neq 1)$. Assume to the contrary that

$$\mathbb{E}(c'(p)|p \neq 1) - w \leq \frac{f(1|e)}{1 - f(1|e)}(c(1) - c'(1)).$$

The ordering $c'(1) < c(1) < w < \mathbb{E}(c'(p)|p \neq 1)$ together with the concavity of u implies

$$u(\mathbb{E}(c'(p)|p \neq 1)) - u(w) < \frac{f(1|e)}{1 - f(1|e)}(u(c(1)) - u(c'(1))).$$

This inequality contradicts the first inequality in (13). Hence,

$$\begin{aligned}&\mathbb{E}(c'(p)|p \neq 1) - w > \frac{f(1|e)}{1 - f(1|e)}(c(1) - c'(1)).\\ \Rightarrow \quad &(1 - f(1|e))\mathbb{E}(c'(p)|p \neq 1) + f(1|e)c'(1) > (1 - f(1|e))w + f(1|e)c(1)\\ \Rightarrow \quad &(1 - f(1|e))\mathbb{E}(c'(p)|p \neq 1) + f(1|e)(c'(1) + \kappa) >\\ &> (1 - f(1|e))w + f(1|e)(c(1) + w - c(1)) = w\end{aligned}$$

because $\kappa > \bar{\kappa} = w - c(1)$. Therefore, contract $\mathcal{W}'$ is more expensive than contract $\mathcal{W}$ in Lemma 5. This contradiction proves that $\beta(p) = 0$ for all $p \in \mathcal{P}$ is optimal for $\kappa > \bar{\kappa}$.

For the second step of the proof, again, compare the principal's costs with and without justification. If she does not provide justification, the principal's costs are w according to Lemma 5. If $e > 0$ and $\kappa < \bar{\kappa} = w - c(1)$, the following contract implies lower costs for the principal:

$$C'(m_P, m_A) = \begin{cases} c(1) & \text{if } m_P = 1 \\ w & \text{else,} \end{cases}$$

$$W'(m_P, m_A) = \begin{cases} c(1) & \text{if } m_P = m_A = 1 \\ w & \text{else} \end{cases}$$

In this contract, the principal justifies the worst evaluation 1. Therefore, $\beta'(1) = 1$ and $\beta'(p) = 0$ for all $p > 1$. The costs of this contract are

$$f(1|e)(c(1)+\kappa)+(1-f(1|e))w < f(1|e)w+(1-f(1|e))w = w.$$

Justifying poor evaluations reduces the principal's costs because justification costs are lower than third-party payments $w - c(1)$. Finally, it remains to verify that contract $\mathcal{W}'$ is feasible.

The agent optimally reports $m_A = \beta$ because his wage in contract $\mathcal{W}'$ is independent of his message. It is easy to verify that $W'(m_P, m_A) \geq C'(m_P, m_A)$ for any combination of messages. Next, consider the principal's reporting strategy. If $p = 1$, reporting $m_P = 1$ implies costs of $c(1) + \kappa$ for the principal. Any deviation increases her costs to at least w. If $p > 1$, reporting $m_P = p$ implies costs of w. Deviating to $\beta = 1$ and a message with $m_P \geq p$ increases her costs to $w + \kappa$. Deviating to $\beta = 1$ and a message with $m_P < p$ is impossible. Deviating to $\beta = 0$ and $m_P \neq p$ weakly increases the principal's costs to w. Therefore, the considered reporting strategy is optimal for the principal. Finally, contract $\mathcal{W}'$ satisfies the agent's participation constraint PC and incentive compatibility IC by the definition of w and $c(1)$. Hence, as long as justification costs are lower than this bound, the principal will justify some evaluations. If $e > 0$ and $\kappa = w - c(1)$, the above steps show that justification and no justification can be optimal. □

Proof of Lemma 7: First, note that the monotone likelihood ratio property ensures

$$\frac{f^H(\delta)}{f^L(\delta)} = \frac{f^H(\delta)}{f^L(\delta)}\frac{\sum_{p'>\delta} f^L(p')}{\sum_{p'>\delta} f^L(p')} < \frac{\sum_{p'>\delta} \frac{f^H(p')}{f^L(p')} f^L(p')}{\sum_{p'>\delta} f^L(p')} = \frac{\sum_{p'>\delta} f^H(p')}{\sum_{p'>\delta} f^L(p')}.$$

Thus, for a given δ merging the performance levels above δ preserves the monotone likelihood ratio property. Hence, according to Grossman and Hart (1983), the strict concavity of $u(\cdot)$ ensures that for a given δ, the solution to Program C is unique. Denote the Lagrange multiplier for the participation constraint by ν_1 and the one for the agent's incentive compatibility by ν_2. Then, Program C yields the first-order conditions:

$$f(p|e) - \nu_1 u'(w(p)) f(p|e) - \nu_2 u'(w(p))(f^H(p) - f^L(p)) = 0 \quad \text{for all } p \leq \delta$$
$$(1 - F(\delta|e)) - \nu_1 u'(w)(1 - F(\delta|e)) - \nu_2 u'(w) \sum_{p>\delta} (f^H(p) - f^L(p)) = 0.$$

Rearranging yields

$$\nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} = \frac{1}{u'(w(p))} \quad \text{for all } p \leq \delta \tag{14}$$

$$\nu_1 + \nu_2 \frac{\sum_{p>\delta} (f^H(p) - f^L(p))}{1 - F(\delta|e)} = \frac{1}{u'(w)} \tag{15}$$

The monotone likelihood ratio property guarantees that

$$\frac{\sum\limits_{p>\delta} (f^H(p) - f^L(p))}{1 - F(\delta|e)} = \frac{\sum\limits_{p>\delta} \frac{f^H(p) - f^L(p)}{f(p|e)} f(p|e)}{\sum_{p>\delta} f(p|e)} > \frac{\frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)} \sum\limits_{p>\delta} f(p|e)}{\sum_{p>\delta} f(p|e)} = \frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)}.$$

Additionally, as before, ν_2 is positive because the solution to Program C without constraint IC is $\delta = 0$ and $w = u^{-1}(\bar{u} + d(e))$ violating IC. Then, (14) implies that $w(p)$ increases in $p \leq \delta$. In addition, equations (14) and (15) as well as the monotone likelihood ratio property imply

$w > w(p)$ for $p \leq \delta$. It remains to verify $w \geq w(\delta) + \kappa$. If $\delta \notin \mathcal{P}$, set $\delta = \max\{p \in \mathcal{P} | p \leq \delta\}$ as a normalization.

Assume, to the contrary, that a solution $\mathcal{W}$ to Program C has $w < w(\delta) + \kappa$. This implies $w(\delta) \in (w - \kappa, w)$ by the preceding steps. Begin with $\delta = 1$. I have already established that ν_2 is positive. Additionally, by standard arguments, ν_1 is positive. Therefore, any solution to Program C must satisfy both constraints PC and IC with equality. For $\delta = 1$, however, the unique solution to constraints PC and IC holding with equality is $w(1) = C(1, 0)$ and $w = C(n, 0)$ with $C(\cdot, \cdot)$ from Lemma 5. By definition of $\bar{\kappa} = w - w(1)$ and $\kappa \leq \bar{\kappa}$, $w(\delta) + \kappa \leq w(\delta) + \bar{\kappa} = w(1) + \bar{\kappa} = w$ contradicting $w < w(\delta) + \kappa$. Consequently, $w < w(\delta) + \kappa$ and $\delta = 1$ is impossible for a solution of Program C.

If $\delta > 1$, modify contract $\mathcal{W}$ by increasing $w(\delta)$ to w and reducing $w(1)$ to $\tilde{w}$, defined by

$$u(\tilde{w}) = u(w(1)) - (u(w) - u(w(\delta)))\frac{f(\delta|e)}{f(1|e)}.$$

The agent's participation constraint PC is satisfied because his expected utility does not change. Additionally, the left-hand side of IC is now larger than the marginal cost of effort $d'(e)$ because

$$\begin{aligned}
&(f^H(1) - f^L(1))(u(\tilde{w}) - u(w(1))) + (f^H(\delta) - f^L(\delta))(u(w) - u(w(\delta))) = \\
=&\frac{f^H(1) - f^L(1)}{f(1|e)} f(1|e)(u(\tilde{w}) - u(w(1))) + \frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)} f(\delta|e)(u(w) - u(w(\delta))) > \\
>&\frac{f^H(1) - f^L(1)}{f(1|e)} \left(f(1|e)(u(\tilde{w}) - u(w(1))) + f(\delta|e)(u(w) - u(w(\delta)))\right) = 0.
\end{aligned}$$

The last equality follows from the definition of $\tilde{w}$. The main inequality follows from the monotone likelihood ratio property. The modified contract satisfies all constraints of Program C and decreases the principal's costs by

$$(w(\delta) + \kappa - w)f(\delta|e) + (w(1) - \tilde{w})f(1|e) > 0.$$

This contradiction shows that $w \geq w(\delta) + \kappa$. The first part of the proof shows that $w(p)$ increases in $p \leq \delta$. Consequently, any solution to Program C satisfies $w \geq w(p) + \kappa$ for all $p \leq \delta$. □

Proof of Proposition 2: In the first step, I characterize the equilibrium behavior for the contract of Proposition 2. The second step shows that this contract is optimal.

Begin with the first step. It is easy to see that $C^*(m_A, m_P) \leq W^*(m_A, m_P)$ for any messages m_A and m_P in the contract of Proposition 2. Suppose that the principal uses the strategy $m_P = p$ and ($\beta(p) = 1$ if and only if $p \leq \delta$). The strategy $m_A = \beta$ following the observed choice β of the principal is optimal for the agent because the agent's wage does not depend on his message m_A. The agent's optimal effort is $e \in \arg\max \sum_{p \leq \delta} u\big(w^{**}(p)\big) f(p|e) + (1 - F(\delta|e))u(w^{**}) - d(e)$, independent of his message. By the same arguments as in Lemma 1, the first-order approach is valid here. Therefore, the agent's incentive compatibility is equivalent to

$$\sum\nolimits_{p \leq \delta} u\big(w^{**}(p)\big)(f^H(p) - f^L(p)) + u(w^{**}) \sum\nolimits_{p \leq \delta} (f^H(p) - f^L(p)) \geq d'(e).$$

Hence, the definition of w^{**}, $w^{**}(p)$ and δ ensures that the desired effort level e is optimal for the agent. The agent accepts the contract if

$$\sum\nolimits_{p\leq\delta} u\big(w^{**}(p)\big)f(p|e) + (1-F(\delta|e))u(w^{**}) - d(e) \geq \bar{u}.$$

Again, the definition of w^{**}, $w^{**}(p)$ and δ guarantees that this condition is satisfied and the agent accepts the contract. To conclude the first step, I show that the contract makes the principal report truthfully $m_P = p$. Suppose that the agent accepts the contract, exerts effort e and sends message $m_A = \beta$. If $p > \delta$, the principal's costs after reporting message m_P are

$$= \begin{cases} w^{**} & \text{if } \beta = 0 \\ w^{**} + \kappa & \text{if } \beta = 1 \end{cases}$$

Hence, if $p > \delta$, there is no profitable deviation for the principal from $\beta = 0$ and $m_P = p$. The case $p \leq \delta$ remains to be considered. Then, the principal's costs after reporting message m_P are

$$\begin{cases} \geq w^{**} & \text{if } \beta = 0 \text{ or } m_P > \delta \\ = w^{**}(m_P) + \kappa & \text{if } \beta = 1 \text{ and } m_P \leq \delta. \end{cases}$$

Lemma 7 ensures $w^{**} \geq w^{**}(p') + \kappa$ for all $p' \in \{1, 2, \ldots, \delta\}$. Remember that $m_P \geq p$ if $\beta = 1$. Therefore, there is no profitable deviation for the principal if $p \leq \delta$. Consequently, truthful reporting $m_P = p$ is optimal for the principal, and the principal's strategy stated above is optimal.

Now turn to the second step. Lemma 4 proves that in any solution to Program B there is a δ, so that $\beta(p) = 1$ if and only if $p \leq \delta$. Lemma 6 ensures that $\delta \geq 1$ for $\kappa < \bar{\kappa}$ and $e > 0$. Lemma 7 proves that the solution to Program B can be determined by analyzing Program C. The first step of this proof shows that the contract in Proposition 2 is feasible and yields equilibrium utilities that solve Program C. Consequently, the contract in Proposition 2 is optimal. □

Proof of Corollary 1: Corollary 1 follows from Proposition 2 and Lemma 6. □

Proof of Proposition 3: Lemma 2 determines a contract $\mathcal{W}$ in the constraint set of Program B with $\delta = n - 1$. Lemma 2 shows that this contract yields lower costs than any contract with $\delta \geq n$. Given $\delta = n - 1$, the definition of $w_e^*(\cdot)$ in Lemma 1 guarantees that contract $\mathcal{W}$ is optimal for sufficiently small costs κ. If $|\{p' \in \mathcal{P} | \beta(p') = 0\}| \geq 2$, Proposition 2 ensures that $\beta(n-1) = \beta(n) = 0$ in an optimal contract. According to Proposition 2, the principal's equilibrium costs satisfy $w(n-1) = w(n)$ in such a contract. Therefore, a feasible contract implies at least wage costs (excluding justification costs) of

$$\begin{aligned} &\min_{w(\cdot)} \sum\nolimits_{p\in\mathcal{P}} w(p)f(p|e) \\ \text{subject to } &\sum\nolimits_{p\in\mathcal{P}} u(w(p))f(p|e) - d(e) \geq \bar{u} && \text{(PC)} \\ &e \in \arg\max \sum\nolimits_{p\in\mathcal{P}} u(w(p))f(p|e) - d(e) && \text{(IC)} \\ &w(n-1) = w(n) && (16) \end{aligned}$$

Lemma 1 proves that a solution to this program without the additional constraint (16) has $w(n-1) < w(n)$. In addition, the solution to this program without the additional constraint (16) is unique according to Grossman and Hart (1983) because $u(\cdot)$ is strictly concave. The idea is to restate the program in terms of utilities. Then, both constraints are linear and the objective function is strictly convex since the inverse function $u^{-1}(\cdot)$ of the strictly concave and increasing utility function $u(\cdot)$ is strictly convex. Hence, the additional constraint is binding, and the wage costs of a contract with $|\{p' \in \mathcal{P} | \beta(p') = 0\}| \geq 2$ are strictly higher than the wage costs of an optimal contract given $\delta = n-1$. The difference between these costs does not depend on κ. Thus, for sufficiently small κ, the additional justification costs κ in the contract given $\delta = n-1$ are lower than the reduction in wage costs. Therefore, $\delta = n-1$ is optimal for a sufficiently small κ. Hence, there is a $\bar{\bar{\kappa}} > 0$ so that $\delta = n-1$ is optimal for $\kappa \leq \bar{\bar{\kappa}}$.

Now consider for a given $\bar{\delta} \in \mathcal{P}$:

$$\begin{aligned} Y(\bar{\delta}) = &\min_{w(\cdot)} \sum\nolimits_{p\in\mathcal{P}} w(p) f(p|e) \\ \text{subject to } &\sum\nolimits_{p\in\mathcal{P}} u(w(p)) f(p|e) - d(e) \geq \bar{u} && \text{(PC)} \\ &e \in \arg\max \sum\nolimits_{p\in\mathcal{P}} u(w(p)) f(p|e) - d(e) && \text{(IC)} \\ &w(p') = w(n) \text{ for all } p' > \bar{\delta} && (17) \end{aligned}$$

The proof of Lemma 7 shows that we can interpret all performance levels above $\bar{\delta}$ as one performance level and the monotone likelihood ratio property is still satisfied. Then, the same arguments as above ensure that the program has a unique solution. As in Lemma 1, the solution $w^{\bar{\delta}}(p)$ to the above program is characterized by

$$\frac{1}{u'(w^{\bar{\delta}}(p))} = \nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} \qquad \forall p \leq \bar{\delta}$$

Therefore, $w^{\bar{\delta}}(p)$ increases in p for all $p \leq \bar{\delta}$. Consequently, comparing $w^{\delta}(\cdot)$ to $w^{\delta+1}(\cdot)$ shows that constraint (17) is binding for $p' = \min\{p \in \mathcal{P} | p > \bar{\delta}\}$. This guarantees $Y(\delta) > Y(\delta+1)$ for $\delta = 1, 2, \ldots, n-2$, because the solution $w^{\delta}(\cdot)$ is unique for each δ. Note that $Y(\bar{\delta})$ does not depend on justification costs κ and that $Y(n-1) = Y(n)$. The principal's total costs are $Y(\bar{\delta}) + F(\bar{\delta}|e)\kappa$. Minimizing this sum with respect to $\bar{\delta} \in \mathcal{P}$ yields an optimal δ that is generically unique.[12] Figure 4 on page 14 depicts the principal's total costs depending on κ. Suppose, to the contrary, that the optimal δ increases in κ, i.e., there are $\kappa, \epsilon > 0$ and $\hat{\delta}, \tilde{\delta} \in \mathcal{P}$ such that $\hat{\delta} < \tilde{\delta}$ and

$$\begin{aligned} Y(\hat{\delta}) + F(\hat{\delta}|e)\kappa &\leq Y(p') + F(p'|e)\kappa && \forall p' \in \mathcal{P} \\ Y(\tilde{\delta}) + F(\tilde{\delta}|e)(\kappa+\epsilon) &\leq Y(p'') + F(p''|e)(\kappa+\epsilon) && \forall p'' \in \mathcal{P} \end{aligned}$$

Set $p' = \tilde{\delta}$ and $p'' = \hat{\delta}$. Then, adding both inequalities yields $F(\tilde{\delta}|e)\epsilon \leq F(\hat{\delta}|e)\epsilon$, contradicting $\hat{\delta} < \tilde{\delta}$. Therefore, no δ' above δ can be optimal if δ is optimal for costs κ and the justification costs increase.

[12] For any open set $K \subset [0, \bar{\kappa}]$ with a continuous distribution, the probability of $\kappa \in K$ with more than one optimal threshold δ is zero.

Finally, denote by $Y(0) = u^{-1}\left(\bar{u} + d(e) + \frac{f(1|e)}{f^L(1)-f^H(1)}d'(e)\right)$ the costs of the optimal contract that does not justify any evaluations following Lemma 6. Remember that $Y(0) = Y(1)+F(1|e)\bar{\kappa}$. Analogously to Lemma 6, it is easy to show that $Y(0) < Y(\delta) + F(\delta|e)\kappa$ for all $\kappa \geq \bar{\kappa}$ and all $\delta \in \{2, \ldots, n\}$. The linearity in κ implies that $\delta = 1$ is optimal for $\kappa \in (\bar{\kappa}-\epsilon, \bar{\kappa}]$ with an $\epsilon > 0$. □

Proof of Corollary 2: For $\kappa > \bar{\bar{\kappa}}$, as defined in Proposition 3, Propositions 2 and 3 as well as Lemma 6 show that the optimal threshold $\delta < n-1$. Hence, wages for several evaluations are pooled as $|\{p \in \mathcal{P}|p > \delta\}| > 1$ because wages for evaluations above δ are constant. Evaluations above δ, however, have a positive variance. Similarly, wages for objective and verifiable performance have a positive variance given a performance above δ according to Lemma 1. Therefore, there is centrality and maximal wage compression at the top. □

Proof of Corollary 3: For $\kappa > \bar{\bar{\kappa}}$ as defined in Proposition 3, Propositions 2 and 3 as well as Lemma 6 show that the optimal threshold $\delta < n-1$. Hence, the principal pools several wages as $|\{p \in \mathcal{P}|p > \delta\}| > 1$. Denote the equilibrium wage in the optimal contract as

$$\hat{c}(p) = \begin{cases} w(p) & \text{if } p \leq \delta \text{ and } \kappa \leq \bar{\kappa} \\ w & \text{if } p > \delta \text{ and } \kappa \leq \bar{\kappa} \\ u^{-1}\left(\bar{u} + d(e) + \frac{f(1|e)d'(e)}{f^L(1)-f^H(1)}\right) & \text{if } p > 1 \text{ and } \kappa > \bar{\kappa} \\ u^{-1}\left(\bar{u} + d(e) - \frac{(1-f(1|e))d'(e)}{f^L(1)-f^H(1)}\right) & \text{if } p = 1 \text{ and } \kappa > \bar{\kappa} \end{cases}$$

for all $p \in \mathcal{P}$ with the values of $w(p)$ and w determined in Program C. Note that in the optimal contracts $\max_{m_P, m_A} C(m_P, m_A) = \hat{c}(n)$. Therefore, optimal contracts imply leniency as

$$\Pr(\{p \in \mathcal{P}|w_e^*(p) = w_e^*(n)\}) = \Pr(p = n) = f(n|e) < \sum_{p'=\max\{2,\delta+1\}}^{n} f(p'|e) = \Pr(\{p \in \mathcal{P}|\hat{c}(p) = \hat{c}(n)\}).$$

Thus, the principal pays the highest wage more often than observing the best performance or paying the highest wage in a setting with a verifiable performance measure. □

Proof of Corollary 4: For a small $\epsilon \geq 0$ and $\alpha \in [0, (1-f^L(1))/(1-f^H(1))]$, define $\bar{f}^H(1) = f^H(1) + \epsilon$ and $\bar{f}^L(1) = f^L(1) + \alpha\epsilon$ while reducing the probability measures $\bar{f}^L(p)$ and $\bar{f}^H(p)$ for better performance levels. For example, $\bar{f}^H(p) = f^H(p) - \epsilon/(|\mathcal{P}|-1)$ and $\bar{f}^L(p) = f^L(p) - \alpha\epsilon/(|\mathcal{P}|-1)$ for all $p > 1$. Suppose ϵ is nonnegative but sufficiently small so that the probability measures $p^L(\cdot)$ and $p^H(\cdot)$ are well defined and the distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property. Note that

$$\frac{\bar{f}(1|e)}{\bar{f}^L(1)-\bar{f}^H(1)} = \frac{f(1|e)+\epsilon(e+\alpha(1-e))}{f^L(1)-f^H(1)-(1-\alpha)\epsilon}.$$

Hence,

$$\partial\frac{\bar{f}(1|e)}{\bar{f}^L(1)-\bar{f}^H(1)}/\partial\epsilon = \frac{(e+\alpha(1-e))(\bar{f}^L(1)-\bar{f}^H(1)) + (1-\alpha)\bar{f}(1|e)}{(\bar{f}^L(1)-\bar{f}^H(1))^2} =$$
$$= \frac{\bar{f}^L(1)-\alpha\bar{f}^H(1)}{(\bar{f}^L(1)-\bar{f}^H(1))^2} > 0$$

because the monotone likelihood ratio property implies that $\bar{f}^L(1) - \bar{f}^H(1) > 0$. Therefore, the principal's payments in the absence of justification increase in ϵ according to Lemma 5. At the same time, the agent's wage for good evaluations increases for $\delta \leq 1$ according to Proposition 2 as well as Lemmas 5 and 6. In addition,

$$\frac{1 - \bar{f}(1|e)}{\bar{f}^L(1) - \bar{f}^H(1)} = \frac{1 - f(1|e) - \epsilon(e + \alpha(1-e))}{f^L(1) - f^H(1) - (1-\alpha)\epsilon}.$$

Hence,

$$\partial \frac{1 - \bar{f}(1|e)}{\bar{f}^L(1) - \bar{f}^H(1)} / \partial \epsilon = \frac{-(e + \alpha(1-e))(\bar{f}^L(1) - \bar{f}^H(1)) + (1-\alpha)(1 - \bar{f}(1|e))}{(\bar{f}^L(1) - \bar{f}^H(1))^2} =$$
$$= \frac{1 - \alpha - \bar{f}^L(1) + \alpha \bar{f}^H(1)}{(\bar{f}^L(1) - \bar{f}^H(1))^2} \geq 0$$

because $\alpha \leq (1 - f^L(1))/(1 - f^H(1))$. Therefore, the agent's wage for the worst evaluation decreases for $\delta \leq 1$ according to Proposition 2 as well as Lemmas 5 and 6. Hence, the threshold $\bar{\kappa}$ increases in ϵ because the utility function $u(\cdot)$ is increasing, so its inverse function is also increasing. Consequently, according to Propositions 2 and 3 as well as Lemma 6, there are some costs κ for which no justifications are used for $\epsilon = 0$, but for small $\epsilon > 0$, the principal offers a contract with justifications. Hence, due to the distribution of costs κ, the probability that the principal uses contracts with justification increases in ϵ. Above, I also showed that such a change implies more variation in wages and payments by the principal. □

Proof of Corollary 5: Suppose $\bar{f}^H(p) = f^H(p)$ and $\bar{f}^L(p) = f^L(p)$ for all $p < n-1$ as well as

$$\bar{f}^H(n-1) = f^H(n-1) - \epsilon, \qquad \bar{f}^H(n) = f^H(n) + \epsilon,$$
$$\bar{f}^L(n-1) = f^L(n-1) - \rho, \qquad \bar{f}^L(n) = f^L(n) + \rho$$

for an $\epsilon \in [-\min\{f^H(n), 1 - f^H(n-1)\}, \min\{f^H(n-1), 1 - f^H(n\}]$
and a $\rho \in [-\min\{f^L(n), 1 - f^L(n-1)\}, \min\{f^L(n-1), 1 - f^L(n\}]$ with
$\epsilon \geq \rho \frac{f^H(n)}{f^L(n)}$, $\epsilon \geq \rho \frac{f^H(n-1)}{f^L(n-1)}$, and $\epsilon < f^H(n-1) - f^L(n-1)\frac{f^H(n-2)}{f^L(n-2)} + \rho \frac{f^H(n-2)}{f^L(n-2)}$. First, note that the constraint set is nonempty because all constraints are satisfied for $\epsilon = \rho = 0$ since the monotone likelihood ratio property ensures that $f^H(n-1) - f^L(n-1)\frac{f^H(n-2)}{f^L(n-2)} > 0$. Indeed, any sufficiently small $\epsilon \geq 0$ satisfies the constraints for $\rho = 0$. Second, $\frac{f^H(n)}{f^L(n)} < \frac{\bar{f}^H(n)}{\bar{f}^L(n)}$ because $\epsilon \geq \rho \frac{f^H(n)}{f^L(n)}$. Third, $\frac{f^H(n-1)}{f^L(n-1)} > \frac{\bar{f}^H(n-1)}{\bar{f}^L(n-1)}$ because $\epsilon \geq \rho \frac{f^H(n-1)}{f^L(n-1)}$. Fourth, the second point and the observation that $\frac{f^H(n-2)}{f^L(n-2)} < \frac{\bar{f}^H(n-1)}{\bar{f}^L(n-1)}$ for $\epsilon < f^H(n-1) - f^L(n-1)\frac{f^H(n-2)}{f^L(n-2)} + \rho \frac{f^H(n-2)}{f^L(n-2)}$ imply that the new distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property. Fifth, ϵ and ρ do not change the expectation of the likelihood ratios $\frac{\bar{f}^H(p) - \bar{f}^L(p)}{\bar{f}(p|e)}$. Hence, the second, third and fifth point imply that the likelihood ratio distribution of $\bar{F}(p|e)$ is a mean-preserving spread compared to the likelihood ratio distribution of $F(p|e)$.

Program C for a given δ satisfies the assumptions of Kim (1995, p. 90) because he writes "our results still hold in a discrete case as well." Kim (1995, Proposition 1) shows that the

mean-preserving spread in the likelihood ratio distribution decreases the principal's wage costs. Therefore, following the proof of Proposition 3, $Y_{\bar{F}(p|e)}(n-1) < Y_{F(p|e)}(n-1)$.

To compute $Y_{\bar{F}(p|e)}(p)$ for any $p < n-1$, consider the likelihood ratio if the two best performance levels are considered together:

$$\frac{\bar{f}^H(n) + \bar{f}^H(n-1) - \bar{f}^L(n) - \bar{f}^L(n-1)}{\bar{f}(n|e) + \bar{f}(n-1|e)} =$$
$$= \frac{f^H(n) + f^H(n-1) - f^L(n) - f^L(n-1)}{f(n|e) + f(n-1|e)}$$

because all ϵ and ρ terms cancel out. Hence, optimal contracts for $\delta \leq n-1$ remain unchanged according to Proposition 2 as well as Lemmas 5 and 6. Hence, the principal's total costs also remain unchanged for a given $\delta \leq n-1$. In the notation of Proposition 3, $Y_{\bar{F}(p|e)}(p) = Y_{F(p|e)}(p)$ for all $p < n-1$. Following the analysis of Proposition 3, there are justification costs for which the optimal δ increases to $n-1$ and the principal justifies strictly more evaluations for the distribution $\bar{F}(p|e)$ than for the distribution $F(p|e)$. For such a κ, the probability of an evaluation with justification increases. Given the distribution of justification costs, the probability of providing justification increases for $\bar{F}(p|e)$ compared to $F(p|e)$. □

Proof of Corollary 6: Suppose there is an $\epsilon > 0$ such that the new disutility of exerting effort $\bar{d}(e) = (1+\epsilon)d(e)$ for all $e \in [0,1)$. Denote by $w_1 = \bar{u} + \bar{d}(e) + \frac{f(1|e)\bar{d}'(e)}{f^L(1) - f^H(1)}$, by $w_0 = \bar{u} + \bar{d}(e) - \bar{d}'(e)\frac{1-f(1|e)}{f^L(1)-f^H(1)}$ and by $(u^{-1})'[w]$ the derivative of the inverse function at w. Then

$$\frac{\partial \bar{\kappa}_{\bar{d}(e)}}{\partial \epsilon} = (u^{-1})'[w_1]\left(d(e) + \frac{f(1|e)d'(e)}{f^L(1) - f^H(1)}\right) - (u^{-1})'[w_0]\left(d(e) - d'(e)\frac{1-f(1|e)}{f^L(1) - f^H(1)}\right) >$$
$$> (u^{-1})'[w_0]\left(d(e) + \frac{f(1|e)d'(e)}{f^L(1) - f^H(1)} - d(e) + d'(e)\frac{1-f(1|e)}{f^L(1) - f^H(1)}\right) =$$
$$= (u^{-1})'[w_0]d'(e)\frac{1}{f^L(1) - f^H(1)} > 0$$

because $w_1 > w_0$ and the utility function $u(\cdot)$ is concave and increasing, so its inverse function is increasing and convex. In addition, the monotone likelihood ratio property implies $f^L(1) - f^H(1) > 0$. For $\delta \leq 1$, Proposition 2 as well as Lemmas 5 and 6 imply that $\bar{\kappa}$ is the difference between the agent's wage for the worst evaluation and good evaluations. Therefore, the variation in the agent's wages increases.

Moreover, according to Propositions 2 and 3 as well as Lemma 6, there are some costs κ for which no justifications are used for $\epsilon = 0$, but for small $\epsilon > 0$, the principal offers a contract with justifications. Hence, due to the distribution of costs κ, the probability that the principal uses contracts with justification increases in ϵ. Above, I also showed that such a change implies more variation in wages and payments by the principal. □

Acknowledgements: Financial support from the German Research Foundation (DFG) through the Collaborative Research Centre (CRC TRR 190 and SFB/TR 15), and from the Leib-

niz Association through Leibniz Science Campus - Berlin Centre for Consumer Policies (BCCP), and Leibniz Competition - Berlin Economics Research Associates (BERA) is gratefully acknowledged.

References

Addison, J. and Belfield, C. R. (2008). The Determinants of Performance Appraisal Systems: A Note (Do Brown and Heywood's Results for Australia Hold Up for Britain?). *British Journal of Industrial Relations*, 46(3):521–531.

Al-Najjar, N. I. and Casadesus-Masanell, R. (2001). Discretion in Agency Contracts. *Harvard Business School Working Paper*, 02–015.

Baker, G., Gibbons, R., and Murphy, K. J. (1994). Subjective Performance Measures in Optimal Incentive Contracts. *Quarterly Journal of Economics*, 109(4):1125–1156.

Bernardin, H. J. and Orban, J. A. (1990). Leniency Effect as a Function of Rating Format, Purpose for Appraisal, and Rater Individual Differences. *Journal of Business and Psychology*, 5(2):197–211.

Bretz, R. D., Milkovich, G. T., and Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, 18(2):321–352.

Brown, M. and Heywood, J. S. (2005). Performance Appraisal Systems: Determinants and Change. *British Journal of Industrial Relations*, 43(4):659–679.

Compte, O. (1998). Communication in Repeated Games with Imperfect Private Monitoring. *Econometrica*, 66(3):597–626.

Dessler, G. (2008). *Human resource management*. Pearson Prentice Hall, 11th edition.

Dewatripont, M. and Tirole, J. (2005). Modes of Communication. *Journal of Political Economy*, 113(6):1217–1238.

Doornik, K. (2010). Incentive contracts with enforcement costs. *Journal of Law, Economics, and Organization*, 26(1):115–143.

Fuchs, W. (2007). Contracting with Repeated Moral Hazard and Private Evaluations. *American Economic Review*, 97(4):1432–1448.

Fuchs, W. (2015). Subjective Evaluations: Discretionary Bonuses and Feedback Credibility. *American Economic Journal: Microeconomics*, 7(1):99–108.

Gale, D. and Hellwig, M. (1985). Incentive-Compatible Debt Contracts: The One-Period Problem. *Review of Economic Studies*, 52(4):647–663.

Gibbons, R. (1998). Incentives in organizations. *Journal of Economic Perspectives*, 12(4):115–132.

Gibbs, M. J., Merchant, K. A., Van Der Stede, W. A., and Vargus, M. E. (2009). Performance Measure Properties and Incentive System Design. *Industrial Relations*, 48(2):237–264.

Giebe, T. and Gürtler, O. (2012). Optimal contracts for lenient supervisors. *Journal of Economic Behavior & Organization*, 81(2):403–420.

Goldluecke, S. and Kranz, S. (2013). Renegotiation-proof relational contracts. *Games and Economic Behavior*, 80:157–178.

Golman, R. and Bhatia, S. (2012). Performance evaluation inflation and compression. *Accounting, Organization and Society*, 37:534–543.

Grossman, S. J. and Hart, O. D. (1983). An analysis of the principal-agent problem. *Econometrica*, 51(1):7–45.

Hall, B. J. and Madigan, C. (2000). Compensation and Performance Evaluation at Arrow Electronics. *Harvard Business School Case*, 800-290.

Hart, O. and Moore, J. (1998). Default and Renegotiation: A Dynamic Model of Debt. *Quarterly Journal of Economics*, 113(1):1–41.

Holmström, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics*, 10(1):74–91.

Jawahar, I. M. and Stone, T. H. (1997). Appraisal Purpose versus Perceived Consequences: The Effects of Appraisal Purpose, Perceived Consequences, and Rater Self-Monitoring on Leniency of Ratings and Decisions. *Research and Practice in Human Resource Management*, 5(1):33–54.

Kambe, S. (2006). Subjective Evaluation in Agency Contracts. *Japanese Economic Review*, 57(1):121–140.

Kampkötter, P. and Sliwka, D. (2018). More Dispersion, Higher Bonuses? On Differentiation in Subjective Performance Evaluations. *Journal of Labor Economics*, 36(2):511–549.

Kim, S. K. (1995). Efficiency of an Information System in an Agency Model. *Econometrica*, 63(1):89–102.

Krasa, S. and Villamil, A. P. (2000). Optimal Contracts when Enforcement Is a Decision Variable. *Econometrica*, 68(1):119–134.

Kvaløy, O. and Olsen, T. E. (2009). Endogenous Verifiability and Relational Contracting. *American Economic Review*, 99(5):2193–2208.

Lang, M. (2018). The Benefits of Stochastic Contracts. *Working Paper*.

Levin, J. (2003). Relational Incentive Contracts. *American Economic Review*, 93(3):835–857.

Li, J. and Matouschek, N. (2013). Managing Conflicts in Relational Contracts. *American Economic Review*, 103(6):2328–2351.

MacLeod, W. B. (2003). Optimal Contracting with Subjective Evaluation. *American Economic Review*, 93(1):216–240.

MacLeod, W. B. (2007). Reputations, Relationships, and Contract Enforcement. *Journal of Economic Literature*, 45(3):595–628.

MacLeod, W. B. and Malcomson, J. M. (1989). Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment. *Econometrica*, 57(2):447–480.

MacLeod, W. B. and Parent, D. (1999). Job characteristics and the form of compensation. *Research in Labor Economics*, 18:177–242.

MacLeod, W. B. and Tan, T. Y. (2017). Opportunism in Principal-Agent Relationships with Subjective Evaluation. *NBER Working Paper*, 22156.

Malcomson, J. M. (2012). Relational Incentive Contracts. In Gibbons, R. and Roberts, J., editors, *Handbook of Organizational Economics*. Princeton University Press. ch. 25:1014–1065.

Murphy, K. J. (1993). Performance measurement and appraisal: Merck tries to motivate managers to do it right. *Employment Relations Today*, 20(1):47–62.

Newman, J. M., Gerhart, B. A., and Milkovich, G. T. (2017). *Compensation*. McGraw-Hill Education, 12th edition.

Noe, R. A., Hollenbeck, J. R., Gerhart, B., and Wright, P. M. (2018). *Human Resource Management: Gaining a Competitive Advantage*. McGraw-Hill, 11th edition.

Pearce, D. G. and Stacchetti, E. (1998). The interaction of implicit and explicit contracts in repeated agency. *Games and Economic Behavior*, 23(1):75–96.

Porter, C., Bingham, C., and Simmonds, D. (2008). *Exploring human resource management*. McGraw-Hill Education, London.

Puhani, P. A. and Yang, P. (2017). Does increased accountability decrease leniency in performance ratings? *Working Paper*.

Rogerson, W. P. (1985). The First-Order Approach to Principal-Agent Problems. *Econometrica*, 53(6):1357–1367.

Rotemberg, J. J. and Saloner, G. (1993). Leadership Style and Incentives. *Management Science*, 39(11):1299–1318.

Spence, J. R. and Keeping, L. (2011). Conscious Rating Distorting in Performance Appraisal: A Review, Commentary, and Proposed Framework for Research. *Human Resource Management Review*, 21:85–95.

Suvorov, A. and van de Ven, J. (2009). Discretionary rewards as a feedback mechanism. *Games and Economic Behavior*, 67(2):665–681.

Taylor, E. K. and Wherry, R. J. (1951). A study of leniency in two rating systems. *Personnel Psychology*, 4(1):39–47.

Townsend, R. M. (1979). Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory*, 21(2):265–293.

Zábojník, J. (2014). Subjective evaluations with performance feedback. *RAND Journal of Economics*, 45(2):341–369.